

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:06:19 ; Search time 2628 seconds  
(without alignments)  
7408.591 Million cell updates/sec

Title: US-09-252-691c-1394  
Perfect score: 669  
Sequence: 1 gctaatgacggcaattgat.....ttacccttaaggagaactga 669

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 20

Total number of hits satisfying chosen parameters: 20

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.st.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23	3.4	10959	1 AE000213	AE000213 Escherich
C 2	23	3.4	11732	1 AE005329	AE005329 Escherich
C 3	23	3.4	15007	1 D90748	D90748 Escherichia
C 4	23	3.4	20284	1 D90749	D90749 Escherichia
C 5	23	3.4	110000	2 AC073744_0	AC073744 Mus muscu
C 6	23	3.4	222605	1 AP002555	AP002555 Escherich
C 7	22	3.3	24578	1 AE008754	AE008754 Salmonell
C 8	22	3.3	254050	1 AL627269	AL627269 Salmonell
C 9	21	3.1	10393	1 AE013675	AE013675 Yersinia
C 10	21	3.1	216050	1 AJ414157	AJ414157 Yersinia
C 11	20	3.0	884	8 AF415022	AF415022 Chamaebat
C 12	20	3.0	7770	1 AB064593	AB064593 Escherich
C 13	20	3.0	10473	1 AE013782	AE013782 Yersinia
C 14	20	3.0	32403	3 U55370	U55370 Caenorhabdi
C 15	20	3.0	137586	9 AC106775	AC106775 Homo sapi
C 16	20	3.0	152037	2 AC128922	AC128922 Rattus no
C 17	20	3.0	171370	9 AC004021	AC004021 Human PAC
C 18	20	3.0	175838	2 AC116356	AC116356 Homo sapi
C 19	20	3.0	193050	1 AJ414149	AJ414149 Yersinia
C 20	20	3.0	195829	9 AL353689	AL353689 Human DNA

ALIGNMENTS

RESULT 1  
AE000213/c  
LOCUS Escherichia coli K12 MG1655 section 103 of 400 of the complete genome. BCT 01-DEC-2000  
DEFINITION Escherichia coli K12 MG1655 section 103 of 400 of the complete genome.  
ACCESSION AE000213 U00096  
VERSION AE000213.1 GI:1787371  
KEYWORDS Escherichia coli K12.  
SOURCE Escherichia coli K12.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 10959)  
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.  
TITLE The complete genome sequence of Escherichia coli K-12  
JOURNAL Science 277 (5331), 1453-1474 (1997)  
MEDLINE 97426617  
PUBMED 9278503  
REFERENCE 2 (bases 1 to 10959)  
AUTHORS Blattner, F.R.  
TITLE Direct Submission  
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459  
REFERENCE 3 (bases 1 to 10959)  
AUTHORS Blattner, F.R.  
TITLE Direct Submission  
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459  
REFERENCE 4 (bases 1 to 10959)  
AUTHORS Plunkett, G. III.  
TITLE Direct Submission  
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

## COMMENT

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

## FEATURES

## source

## Location/Qualifiers

1..10959 /organism="Escherichia coli K12"  
/strain="K12"  
/sub\_strain="MG1655"  
13..40 /db\_xref="taxon:83333"  
13..40  
/note="factor Sigma70; predicted +1 start at 1184959"  
119..136  
/note="central position to predicted promoter:82"  
/bound\_molety="PhoB predicted site"  
155..1381  
/gene="pept"  
/note="b1127"  
155..1381  
/gene="pept"  
/EC\_number="3.4.11.-"  
/function="putative enzyme; Not classified"  
/note="o408; 100 pct identical to 42 aa fragment  
PEPT\_ECOLISW: P29745; 93 pct identical to 407 residues of  
409 aa PEPT\_SALTY SW: P26311"  
/codon\_start=1  
/transl\_table=11  
/product="putative peptidase T"  
/protein\_id="AAC74211.1"  
/db\_xref="GI:1787372"

## promoter

## protein\_bind

## gene

## CDS

## promoter

## gene

## CDS

/protein\_id="AAC74212.1"  
/db\_xref="GI:1787373"  
/translation="MLNMEYQLTNWPDFLERHWKRPVVLKRGNNFIDPISPDELA  
GLAMESVDRLSHODGKQVSHGPFESYDHLEGTNKSLLVOAVNHWHPTALMRP  
FRELPHRRIDDLMISSFPVGGVGHLDQVDVFIQTGRRRVGKGLQKHCCHP  
DLQVDFEALIDEELEPGLILYPGFPHEGYALEMANNYSVGFAPNAPRELISGFA  
DYVLQRELGNYSDDPPYRAHPADYLPQEMDKLREMLLELINQPHFKQWFEFTS  
QSRHLEIDAPPEPPYQDEYDALKQCEVLVRLGGLRVLRIGDDVYANGEKIDSPHRP  
ALDALSNIALTAENFGDALEDPSEFLAALVNSGVWFEEG"  
complement(2583..2613)  
/note="factor Sigma70; predicted +1 start at 1187488"  
complement(2627..4087)  
/gene="phoQ"  
/note="b1129"  
complement(2627..4087)  
/gene="phoQ"  
/EC\_number="2.7.3.-"  
/function="enzyme; Global regulatory functions"  
/note="f486; 99 pct identical to PHOQ\_ECOLI SW: P23837"  
/codon\_start=1  
/transl\_table=11  
/product="sensor protein PhoQ"  
/protein\_id="AAC74213.1"  
/db\_xref="GI:1787374"  
/translation="MRKLLRFFLSLRVRLTAATAVVLVSLAYGMVALIGYSVSF  
DKTTFLLRGESNLFYTLAKWENKHLVELPENIDKQSTMTLIYDENGQLLWQARDV  
PWLKMTQDPLKSNGFHEADVNDTSLLSGDHSTQQOQLQVREDDDAETHSVA  
VNYVPATSRMPKLTIVVDTPYVELKSSYMWKFIYVLSANLLVTPLLVAHWNSL  
RPIEALKEVRELEHNRLLNATRELISLRNLRLKLSEREDYDKTTRTLDLT  
HSLTAPLAVLOSTLSLRSEKMSVSDAEPMLEOISQOIGYILHRASMRGTLTLLS  
RELHPVAPLDLNTSALNKVYQRKVNISLIDISPEISVGEQNDFVEMGNVLDNACK  
YCLEFVEISARQTDHLYIVVEDDGPGLPSKREIVFDGRQVRDTPRGQGVGLAVAR  
EITEVEGKIVAGESMLGARMVEIFGRQHSAPKDE"  
complement(4087..4758)  
/gene="phoP"  
/note="b1130"  
complement(4087..4758)  
/gene="phoP"  
/function="regulator; Global regulatory functions"  
/note="f223; 100 pct identical to PHOP\_ECOLI SW: P23836"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulatory protein"  
/protein\_id="AAC74214.1"  
/db\_xref="GI:1787375"  
/translation="MRVLVVEDNALLRHLLKVOIQDAGHOVDDADEAKEADYVLENEHI  
PDIAIVDLGLPDEGLSIRWRSDVSLPILVLTARESQDKVELVLSACADYVTKP  
FHIEEVMARQALMRNSGLASQVISLPFPQVLSRRELINDEVIKLTAFETIMET  
LIRNNGKVSQKSLMLQLYPDAELRESHTIDVLMGLURKKIKQAQYPOEVITTVRGQSY  
LFELR"  
complement(4257..4287)  
/gene="phoP"  
/note="factor Sigma70; predicted +1 start at 1189162"  
complement(4927..6297)  
/gene="purB"  
/note="b1131"  
complement(4927..6297)  
/gene="purB"  
/EC\_number="4.3.2.2"  
/function="enzyme; Purine ribonucleotide biosynthesis"  
/note="f456; 99 pct identical to PURB\_ECOLI SW: P25739"  
/codon\_start=1  
/transl\_table=11  
/product="adenylosuccinate lyase"  
/protein\_id="AAC74215.1"  
/db\_xref="GI:1787376"  
/translation="MELSSLTAVSPVDGRYGDVKSALRGIFSEYGLKFRVQVEVRWL  
OKLAAHAAIKVEPAFAADAIGYLDIAIVASPEEDAARIKTERTNHDVKAVEFLKE  
KVAIEIPELHAVSEFIHFACSEDINLSHALMLKLTARDEVILPYWROLIDGINKDLAVQ  
YRDIPLLSRTHQGPATPSTIGKEMANVAYRMERQYRQLNQVEILGKINGAVGNNAHI  
AAYPEVDWHQFSEEFVTSLGIQWNPTTQLEPHDYIAELFDCAVFTKILIDFDROVW  
GYIALNHFQKRTAGESTMPKVPIDFENSEGNLSLNAVLSKSLKPLPSRWQ  
RLDSTDVLRNLGVGIGYALIAVQSTLKGYSKLEVNDRDLHLLDHNWEVLAEPIQTV

## promoter

## gene

## CDS

```

gene      MRRYIEKPYEKLKELTRGKRVDAEGMKQFDIGLALPEEEKARLKAMTPANYIGRAIT
          MVDELK"
          complement(6301..6942)
          /gene="ycfC"
          /note="b1132"
          complement(6301..6942)
          /gene="ycfC"
          /functions="orf; Unknown"
          /note="f213; 99 pct identical to YCF_CECOLI SW: P25746"
          /codon_start=1
          /transl_table=11
          /product="orf, hypothetical protein"
          /protein_id="AAC74216.1"
          /db_xref="GI:1787377"
          /translation="MAKNYYDITLALAGIQCASRLVOOLAHQHCADADALHVSLSNII
          DMPSTLAVFGGSEANLRVGLTLLGVNLASSRQGLNAELTYTSLMVLKRLSSA
          3.48; Score 23; DB 1; Length 10959;
          Best Local Similarity 100.0%; Pred. No. 0.19;
          Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 133 CTGTTCAATAAACCTAGGATGT 155
      |||||
Db 9136 CTGTTCAATAAACCTAGGATGT-9114

```

77

```

RESULT 2
AE005329/c 11732 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 153
of 155.
ACCESSION AE005329
VERSION AE005329.1 GI:12514784
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE
1 (bases 1 to 11732)
  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
  Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
  Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
  Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoukis,K.,
  Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
  Welch,R.A. and Blattner,F.R.
  Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  Nature 409 (6819), 529-533 (2001)
11206551
2 (bases 1 to 11732)
  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
  Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
  Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
  Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoukis,K.,
  Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
  Welch,R.A. and Blattner,F.R.
  Direct Submission
  Submitted (22-OCT-2000) Laboratory of Genetics, University of
  Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
  Location/Qualifiers
    1. .11732
      /organism="Escherichia coli O157:H7 EDL933"
      /strain="EDL933"
      /serotype="O157:H7"
      /db_xref="taxon:155864"
      /note="enterohemorrhagic"
      complement(115..1485)
      /gene="purB"
      /note="Z1860"
      complement(115..1485)
      /gene="purB"
      /functions="enzyme; Nucleotide biosynthesis; Purine
      ribonucleotide biosynthesis"

```

```

/note="Residues 1 to 456 of 456 are 99.78 pct identical to
residues 1 to 456 of 456 from Escherichia coli K-12 Strain
MG1655: B1131"
/codon_start=1
/transl_table=11
/product="adenylosuccinate lyase"
/protein_id="AAG55957.1"
/db_xref="GI:12514785"
/translation="MELSSLTAVSPVDGRYKVSALRGIFSEYGLLKFRVQVEVRWL
OKIAHAHAIKEVFAFADAIGYDAIVANSEEDAAIKTERTTNDHVKAVEFLKE
KVAIEPILSRTHGQPATPSTIGKMANVAYRMEQYROLNOVEILGKINGAVGNNAHI
AYPEVDWHQFSEEFVTSIGIOWNPYTQIEPHDYIAELFDCVARENTIILIDEDRDYH
GYALNHFKAKTTAGEIGSSTMPHKVNPIDFENSEGNLGSNAVLQHLAKLPSVRNQ
RDUTDSTVLNKLGVIGYALGYATSLGKLVSKLEVRNDRDLHLDHNEVLAEPITQV
MRYGIEKPYEKLKELTRGKRVDAEGMKQFDIGLALPEEEKARLKAMTPANYIGRAIT
MVDELK"
complement(1489..2130)
/gene="ycfC"
/note="Z1861"
complement(1489..2130)
/gene="ycfC"
/function="orf; Unknown function"
/note="Residues 1 to 213 of 213 are 99.53 pct identical to
residues 1 to 213 of 213 from Escherichia coli K-12 Strain
MG1655: B1132"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55958.1"
/db_xref="GI:12514786"
/translation="MAKNYYDITLALAGIQCASRLVOOLAHQHCADADALHVSLSNII
DMPSTLAVFGGSEANLRVGLTLLGVNLASSRQGLNAELTYTSLMVLKRLSSA
KQALDTLGNRLNGIQLQLEHFLDQSLMSAAIYVDVISPGLQVIGTQVGGSPAVLOS
POQAKYVRATLLAGIRAAVHLHQVGGGRLQLMFSRNRLTTQAKQILAHLP"
complement(2166..3317)
/gene="ycfB"
/note="Z1862"
complement(2166..3317)
/gene="ycfB"
/function="orf; Unknown function"
/note="Residues 1 to 383 of 383 are 99.47 pct identical to
residues 1 to 383 of 383 from Escherichia coli K-12 Strain
MG1655: B1133"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55959.1"
/db_xref="GI:12514787"
/translation="MIEYAALKENYVSDPMSETAKKVIIVGMSGVDSVSNWLLQOOG
YQVEGLFMKNWEEDGEEYCTAAADLAQAVCDKIGIELHTVNFRAEYWDNVEFLFL
AEYKAGTPNPDIILCNKEIKFAPLEADGADYIATGYVRRADVDDKSRLLRGL
DSNKDSYFLYTLISHEQIAQSLFPGVELEKQVQRKIAEDGLVTAKKDGTGTCFGE
RKRFEGRYLPAQPGKIITVDGDEIGEHOGLMYHTLQGRKGLIGTGTGKTEPWTY
TKVDVNNILIVAGQHEPRLMSVGLIAQOLHWVDREPFTGTMRCTVKTTRQTDIPC
TVKALDDRIEIVFDEPVAATPCQSAVFNGEYVCLGGGIIIEORLPLI"
complement(3326..3787)
/gene="Z1863"
complement(3326..3787)
/gene="Z1863"
/function="putative enzyme; Not classified"
/note="Residues 1 to 153 of 153 are 100.00 pct identical
to residues 1 to 153 of 153 from Escherichia coli K-12
Strain MG1655: B1134"
/codon_start=1
/transl_table=11
/product="putative phosphohydrolase"
/protein_id="AAG55960.1"
/db_xref="GI:12514788"
/translation="MFKPHVTACVHHAEGKFLVVEETINGKALWNQAPAGHLEADEFL
VEAAARELWEETGISAQPHFIRMHOWIAPDKTFLFLFAIELEQICQPPHDSID
CCRWVSAEIILOASNLRLPLVAESIRCYQSQRYPLEMIGDFNWPFTKGI"
complement(3797..4420)

```

```

/ gene="ymfC"
/ note="Z1864"
/ complement(3797..4420)
/ gene="ymfC"
/ function="orf; Unknown function"
/ notes="Residues 1 to 207 of 207 are 99.03 pct identical to
residues 1 to 207 of 207 from Escherichia coli K-12 Strain
MG1655: B1135"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAG55961.1"
/ db_xref="GI:12514789"
/ translation="MKTSTFRNQVRFSSQSTRRRKPNQTRVILFNKPYDVLQPF
TDAAGRTKLEFIPVGVGAAGRLDRDSEGLVLTNNGALQARLTQPGRTGRIYVQ
VEGIPQDALEALNRGVTLDNPTLPAGAEVDEPAWLWPNPPIREKRSIPTSWLKI
TLVEGRNROVRRTAHVGPFTLLIRYANGDYSLDNLANGWREATE"
4622..5872
/ gene="icdA"
/ note="Z1865"
4622..5872
/ gene="icdA"
/ function="enzyme; Energy metabolism, carbon: TCA cycle"
/ notes="Residues 1 to 416 of 416 are 99.27 pct identical to
residues 1 to 416 of 416 from Escherichia coli K-12 Strain
MG1655: B1136"
/ codon_start=1
/ transl_table=11
/ product="isocitrate dehydrogenase, specific for NADP+"
/ protein_id="AAG55962.1"
/ db_xref="GI:12514790"
/ translation="MKSIVVPAQKKITLQNKLVNPNPIPYIEGGIGVDVTPA
MLKVDAAEVAKYNGKRSIMWEIYTGKSTQYGVQDVMVLPAPETLIRVRAIRGP
LTPVGGIRSLRVKALQELDYICLRPVYIQGTSPVKKHPELTDVIFRENSDIY
AGLEWASDAEKVKIFLREMGVKKIRFPEHCIGIKPCSEEGKRLVRAAIEVAI
ANDRDSVTLVHKINIMKTEGAFKDWGLYELAREFGGELIDGGPWLKVNPTGKEIV
IKDYADAFLQQLLPAEDYVYACNLDYISDALAAQVGGIGTAPAGNIGDEAL
FEATHTGAKLQCKVNGFSIILSAEMMLRMGMWTEADLIYKMGEGAINAKTVTD
FERLMGAKLQCKSEFGDAIKNM"
5791..>11732
/ note="O-island #52; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
/ complement(5986..7128)
/ gene="Z1866"
/ gene="Z1866"
/ function="putative enzyme; Integration, recombination
(Phase or Prophage Related)"
/ notes="Residues 1 to 380 of 380 are 98.42 pct identical to
residues 1 to 380 of 380 from Genpept 118 :
9112154501gb|AAA32336.1| (M61865) integrase [Bacteriophage
21]"
/ codon_start=1
/ transl_table=11
/ product="putative integrase of prophage CP-933X"
/ protein_id="AAG55963.1"
/ db_xref="GI:12514791"
/ translation="MAASPRSHKISIPNLYCKDKRTKGVYQYKPLSGRHSLGTD
ENKAKVATEANTYIAEQRTROILSVNERLKMGRSDITVTEWLDKYNISQEDRLQ
HNELPNSYRQKGPRLPREHCGMOHLKDITALDIAELIDAVKAEHNRMAQVPMV
LIDVFEAGHAGVHPGFPNPAQTKQPRNVRNQRSLSEWQAIFDSVSRQPYLKCGL
MLCALVTGRLSDICLKFSDIWDMLHITQETKSKLAIPLNLCDAINITLREVIS
QCDADVSYLYHYRHTTSQANGQVSANTLTAFKAKREKCGIKWEGQTAPTFHEQ
RSLSLRYREQGLDTOKLLGHKSRKMTDRYNDNRGDKIIVDIKTA"
complement(7118..7354)
/ gene="Z1867"
/ complement(7118..7354)
/ gene="Z1867"
/ function="putative enzyme; Integration, recombination
(Phase or Prophage Related)"
/ notes="Residues 1 to 78 of 78 are 98.71 pct identical to
residues 1 to 78 of 78 from Genpept 118 :
9112154491gb|AAA32335.1| (M61865) excisionase

```

```

[ Bacteriophage 21 ]
/codon_start=1
/transl_table=11
/product="putative integrase of prophage CP-933X"
/db_xref="GI:12514792"
/translation="MSRLITLDRWAKEEFGDLAPSRVLKKAQGMMAIPAIVGVR
WMIDRNSREVGTLAEPQLPINAPKLQRIADGC"
7458..8201

Query Match      3.4%; Score 23; DB 1; Length 11732;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGTTCAATAAACCTACGATGT 155
      |||||
DB 4324 CTGTTCAATAAACCTACGATGT 4302

RESULT 3
D90748/c
LOCUS      D90748      15007 bp      DNA      linear      BCT 07-FEB-1999
DEFINITION Escherichia coli genomic DNA. (25.6 - 25.9 min).
ACCESSION  D90748 AB001340
VERSION    D90748.1 GI:1651553
KEYWORDS   Complete and shotgun sequencing; potB; potA; pepT; phoQ; phoP;
           purB; ycfC; ycfB; icdA; icd; icdE; lit.
SOURCE     Escherichia coli (strain:K12) DNA, clone:Kohara clone #239.
ORGANISM   Escherichia coli
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia
           1 (sites)
REFERENCE  Oshima, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
           Oshimoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
           Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
           Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
           Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K.,
           Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T.
           A 718-kb DNA sequence of the Escherichia coli K-12 genome
           corresponding to the 12.7-28.0 min region on the linkage map
           DNA Res. 3 (3), 137-155 (1996)
           97061202
           2 (sites)
           Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T.,
           Ikemoto, K., Inada, T., Isono, K., Kimura, S., Kimura, S., Kitagawa, M.,
           Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
           Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
           Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
           Yamamoto, Y. and Yano, M.
           The systematic sequencing of the Escherichia coli genome in Japan
           Unpublished
           3 (bases 1 to 15007)
           Mori, H.
Direct Submission
Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
and Technology, Res. & Edu. Center for Genetic Info.: 8916-5
Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@ict.aist-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669)
Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,
Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,

```

Yamamoto,Y. and Yano,M.

Headed by:

Name: Takashi Horiuchi

Address: National Institute of Basic Biology, Okazaki, 444, Japan

E-mail: kishori@nibb.ac.jp

Information operator:

Name: Hirotada Mori

Address: NARA Institute of Science and Technology,

Ikoma, 630-01, Japan

E-mail: hmori@gtc.aist-nara.ac.jp

URL:

The Japan E. coli genome database

http://bsw3.aist-nara.ac.jp.

Location/Qualifiers

1..15007

/organism="Escherichia coli"

/strain="K12"

/db\_xref="taxon:562"

/map="25.6-25.9 min"

/clone="Kohara clone #239"

/note="Nucleotide position 1185468-1200474 from the

initiation site of *thraA* (0 min.). This clone is from

Kohara lambda miniset library."

complement(1..678)

/gene="potB"

complement(<1..678)

/gene="potB"

/note="ORF\_ID:0238#14

similar to SwissProt Accession Number P23860"

/codon\_start=1

/transl\_table=11

/product="Spermidine/putrescine transport system permease

protein PotB."

/protein\_id="BAA35947.1"

/db\_xref="GI:1651554"

/translation="MIVTIVGWLVLVFLPNLMIITGTSFLTRDDASFVKMVFLLDNYT

RLDPLFEVLHSLNLMALIALACVLGYPFAWFLAKLPHKVRFLLLFLIVPWTN

SLIRYGLKIFLSTYKGLNEFLWGLVIDTPIRIMFTPSAVTIGLVILLPFWNPLY

SSIEKDKPLLEAARDLGASKLQTFIRIIPLTMPGIAGCLLVMLPAMGLFYVSDLM

GGAKNLI."

complement(692..1828)

/gene="potA"

complement(692..1828)

/gene="potA"

/note="ORF\_ID:0239#1

similar to PIR Accession Number A40840"

/codon\_start=1

/transl\_table=11

/product="Spermidine/putrescine transport protein A"

/protein\_id="BAA35948.1"

/db\_xref="GI:1651555"

/translation="MGQSKLNKQPSSLSPLVLQAGIRKCFDGEVTPQLDLTINNGE

FLTLGPGSGKTVLRLLIAGLETVDSDRIMLDNEDITHVPAENRYVNTVFQSYALFP

HMTVFVGLRGMOKTPAAETPRVMEALRMVQLETFAQRKPOLSGGQQQRVAIAR

AVANKPRLLLDSELDYKLRQMOKELAKORKLGITFVFVTHDQEALTWSDRIC

VVMRGRIEODGTREIYEKPNLFIAGFGEINNFNATVIERLDEQRVANVEGREG

NIVYNFAVEPGKHLVLRLEDLVRVEINDDNHAEGLIGYVRERYKGMTLESVVELE

NGKMWVSEFNEDDPDFDHSLDQKMAINWVESVEVLADEEHK"

2078..3304

/gene="pept"

2078..3304

/gene="pept"

/note="ORF\_ID:0239#2

similar to SwissProt Accession Number P29745"

/codon\_start=1

/transl\_table=11

/product="Peptidase T (EC 3.4.11.-) (aminotripeptidase)

(tripeptidase) "

/protein\_id="BAA35949.1"

/db\_xref="GI:1651556"

/translation="MDKLRFNLVSLDTQSKAGVRQVPSTGEQKWLHLLKQLEE

MGLNVLTSERKGMATLPANVPDIPAIIGTISHVDTPSDCSGKNVPQIVENTRGGD

IALTGDVLSVPVMPFVLLQLGQTLITTDCKTLTGADDDKAGIAEIMTALAVIQKKI

PHGDIRVAFTPDEEVGKGAKHFDVDAFDARWAYTVDGGGVGELEFENFNAASVNIKIV  
GNNVHGPTAKGVNALSIAARIHAEVPADESPENTEGYEGYHLASMKGTVERADMH  
YIIRFDRKQFARRKRMKEIAKKVKGKGLHPDCYIELVIEDTSYNNRMKEVWHPHLLD  
IAQAMRDCCIIEPLKPIRGGTGGAQLSPFMGLPCPNLFTGGYNYHGKHBEFVLEGMEX  
AVQVIVIRIAELTAQRK"

complement(join(3353..4474,4550..6010))

/gene="phoQ"

complement(3353..4474)

/gene="phoQ"

/note="ORF\_ID:0239#3

similar to PIR Accession Number D41966"

/codon\_start=1

/transl\_table=11

/product="Orf 2 downstream of phoQ."

/protein\_id="BAA35950.1"

/db\_xref="GI:4062695"

/translation="MEYQLTLNMPDFLERHWKRPVVLKRGNNFTIDPISPDLAGLA

MESEYDSRLVSHQDGKQWVSHGPFESYDHLGTFNMSLLVQAVNHHHEPTAALMRPFRE

LPDMRIDDLMISSFVPGGVGPHLDQYDFVFIIGTGRRRRVGKELQMKQKHPHDL

QVDPFEAIIIDELEPGDILYIPPGFPHGCVALENAMNYSVGFRAINTRRELISGFADYV

LORELGGNYSDPDVPPRAHPADVLQPMDKLREMLLELINOPEHFKWQGFISQSR

HELDIAPPEPPQDPEIYDALQGEVLVRLGGLRVLRIGDDVIANGEKIDSPHRPALLD

ALASNIALTAEFNFGDALEDPSEFLAMLAALVNSGYWFFEG"

complement(4550..6010)

/gene="phoQ"

/note="ORF\_ID:0239#4

similar to PIR Accession Number B41966"

/codon\_start=1

/transl\_table=11

/product="Virulence membrane protein phoQ."

/protein\_id="BAA35951.1"

/db\_xref="GI:1651557"

/translation="MKKLLRFFLSLRVRFLLATAVVLVLSLAYGVALIGYSVSF

DKTFRLLRGESNLFYTLAKWNNKLHVELPENIDKQSPMTFLIYDENGQLLWAAQRDV

PWLKMIQDFDLKNSGFHEIADVNDTSLLSGDSHSIQQLQOEVEDDDDAEMTHSVA

RVYATATSRMPKLTIIVVDITIPVELKSSYMWMSFYVLNLANLLVPLLVAAWMSL

NPFEALAKEVELEHNRELLNPATRELTSLVRNLRLLKSERERYDKYRTTLDLT

HSLKPLAVLOSTLRSLEKMSVSDAEPVLEQISRTSQOIGYVLHRASMRGGTLLS

RELHPVLONLTSALNKVQKGVNISLDSPEISFSGEONDVEYMGNVLDNACK

YCLEFVEISAKQTDHEHLYIVEDDGGPILSKREVIFDRGQRVDTLRPGQGVGLAVAR

EITEQYEGKIVAGESMLGARMVIFGRQHSAPKDE"

complement(6010..6681)

/gene="phop"

complement(6010..6681)

/gene="phop"

/note="ORF\_ID:0240#1

similar to SwissProt Accession Number P23836"

/codon\_start=1

/transl\_table=11

/product="transcriptional regulatory protein Phop."

/protein\_id="BAA35952.1"

/db\_xref="GI:1651558"

/translation="MRVLVEDNALLRHLHKVQIQDAGHQVDDAEDAEADYLLNEHI

PDIAIVGLGDPEDGLSLIRWRNSDVSLPILVLTARESQQKVEVLSAGADYVTKP

PHIERVMARMAALMRNSGLASOVLSLPFQVDSLRRRLSINDEVIKLTAFETVIMET

LIRNKGKVVSKDSLMLQLYPDALRESHTIDVLMGLRKKIKQAQVQEVITVRGQGY

LFELR"

complement(6850..8220)

/gene="purB"

complement(6850..8220)

/gene="purB"

/note="ORF\_ID:0240#2

similar to PIR Accession Number S19212"

/codon\_start=1

/transl\_table=11

/product="Adenylosuccinate lyase (EC 4.3.2.2)"

/protein\_id="BAA35953.1"

/db\_xref="GI:1651559"

/translation="MELSSLTAVSPVDGRYGDVSKVLSALRGIFSEYGLLFRVQVEVRWL

QKLAHAHAKEVPAPFAADIGYLVASVFSEEDAAIKTIERTTNHVDYKAVEYFLKE

KVAETPELPGVSEFTFACSTEDINNLSHALMKTARDEVILPYNLDYDKGLDAVQ

YRDIPLLSRTHGQPATPTIGKEMANVAYRMEROYRQLNQVEILGKINGAGVNYNAHI

AAYPEVDNHQSEEVTSIGTQWNPYTTQIEPHDYIAELFDCVARENTLILDFDRDV



/transl\_table=11  
/product="Hypothetical protein 23."  
/protein\_id="BAA35963.1"  
/db\_xref="GI:4062714"  
/translation="MARNYDITLALAGICQARLVQOALHQHCHDADALHVSLSNII  
DMNPSSLAIVFGSEANLVGLLETLGLVFNASSROGLNAELTRYTLTLMLERKLSSA  
KALDITLGNRIHQLEHFDLQSETLMSAMAAIYVDVTSPLGRIVTQSGSPAVLOS  
POVQAKVRATLACIRAAVILWHQVGGGRLQLMFNRRLTTQAKILAHLPTEL"  
complement(3218..4324)  
/gene="ycfB"  
/complement(3218..4324)  
/note="ycfB"  
/gene="ycfB"  
/note="ORF\_ID:0240#4  
similar to SwissProt Accession Number P25745"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein in purB 5' region (orf-15)"  
/protein\_id="BAA35964.1"  
/db\_xref="GI:4062715"  
/translation="MSETAKVIIVGMSGVDSSVAMLLQOQGVQVEGLFMKNWEEDD  
GEECTAAADLAQAQCDKLGIELHTVNFNAAEYWDNVFELFVLAERKGRTPNPDLIC  
NKEIKFALEFAEDLGDGIATGHVVRADVDGKRLRLGLSDNKDQSFLLTSLH  
EQIAQSLFVPELEKPPQVKTAEGLVLTAKKDKSTGICFCIGERKFRFGLRILPAQ  
GKIITVDDEIGEHQGLMYHTLQKRGKLGIGTKEGTEEPWVYVVDKVDVNNILVPAQ  
HEHRLMSVGLIAQLHWVDEPFTGMTCTVKTQYRQTDIPCTVKTALDDDDRIEVED  
EPVAAVTPGQSAVFYNGEVCILGGGIIQRULPLV"  
complement(4378..4839)  
/note="ORF\_ID:0240#5  
similar to PIR Accession Number S64074"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein YGL067"  
/protein\_id="BAA35965.1"  
/db\_xref="GI:4062716"  
/translation="MFKPHVTVCVHAEKGLVVEETINGKALWNPAGHLEADETL  
VEAAALEWETGISAQOPEFVHOMWIAPTKPTFLFLFALELEQICPTOPHSDSID  
CCRWSAEELIQASNLRSPLVAESIRCYSQSQRYPLEMIGDNFFPKGVI"  
complement(4849..5502)  
/note="ORF\_ID:0240#6  
similar to PIR Accession Number I64156"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein HI0694"  
/protein\_id="BAA35966.1"  
/db\_xref="GI:4062717"  
/translation="MRQFIISENTMOKTSFRNHQVRFSSQRSTRKRPENQPTRVILF  
NKPVDVLPQFDEAGRTLKKEFIPQGVYAAAGRLDRDSEGLLVLTNNGALQARLTQPG  
KRCKIYVQVEGIPDTDALEALRNGVTLNDGPTLPAGAEVLDPALWPRNPPIRER  
KSIPTSMLKILYEGNRNQVRMTAHVGFPLRLIRVAMGDYSLDNLANGEWEVTD"  
5674..6924  
/gene="icd"  
/note="icdA; icdE"  
5674..6924  
/gene="icd"  
/note="ORF\_ID:0240#7  
similar to SwissProt Accession Number P08200"  
/codon\_start=1  
/transl\_table=11  
/product="Isocitrate dehydrogenase (NADP) (EC 1.1.1.42)  
(Oxalosuccinate decarboxylase) (icd) (NADP-specific icdh)  
(icd)."  
/protein\_id="BAA35967.1"  
/db\_xref="GI:1651566"  
/translation="MESKVVVPAQCKITLQKLNVPENPIPIYIEGDGIGVDVTPA  
MLKVDAVERKAYGERKISWMEIVTGEKTVQGVQDWLPAETLDLIREYVAKPG  
LITPVGGSIRSLNALQELDLVLCLEPRVYQCTPSPVKHPELTDVIFRENSDIY  
AGIEWKSDSAERKIVFLREEMGVKIRFFPHCGIGKPCSEBGTKKLVRAAEIAY

Query Match 3.4%; Score 23; DB 1; Length 20284;  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 CTGTTCAATAAACCTACGATGT 155  
|||||  
Db 5376 CTGTTCAATAAACCTACGATGT 5354  
|||||  
RESULT 5  
AC073744.0/c  
WPCOMMENT  
Sequence split into 5 fragments LOCUS AC073744 Accession AC073744  
Fragment Name Begin End  
AC073744\_0 1 110000  
AC073744\_1 100001 210000  
AC073744\_2 200001 310000  
AC073744\_3 300001 410000  
AC073744\_4 400001 466859  
LOCUS AC073744 466859 bp DNA linear HTG 12-JUL-2000  
DEFINITION Mus musculus clone RP23-288A3, WORKING DRAFT SEQUENCE, 196  
unordered pieces.  
ACCESSION AC073744  
VERSION AC073744.1 GI:8810361  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 466859)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 466859)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 1847467  
Center clone name: RPCI-23\_288A3  
-----  
Summary Statistics  
Consensus quality: 310234 bases at least Q40  
Consensus quality: 387412 bases at least Q30  
Consensus quality: 406374 bases at least Q20  
Estimated insert size: 193000; pulse field gel estimation  
Estimated insert size: 447359; sum-of-contigs estimation  
Quality coverage: 6.42 in Q20 bases; pulse field gel estimation  
Quality coverage: 2.77 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 196 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1152: contig of 1152 bp in length  
\* 1153 1252: gap of unknown length  
\* 1253 2396: contig of 1144 bp in length  
\* 2397 2496: gap of unknown length  
\* 2497 4003: contig of 1507 bp in length  
\* 4004 4103: gap of unknown length  
\* 4104 5282: contig of 1179 bp in length  
\* 5283 5382: gap of unknown length  
\* 5383 6463: contig of 1081 bp in length  
\* 6464 6564: gap of unknown length  
\* 6564 7611: contig of 1048 bp in length  
\* 7612 7711: gap of unknown length  
\* 7712 9030: contig of 1319 bp in length  
\* 9031 9130: gap of unknown length

9131	10349:	contig of 1219	bp in length
10350	10449:	gap of unknown length	
10450	11554:	contig of 1105	bp in length
11555	11654:	gap of unknown length	
11555	12992:	contig of 1338	bp in length
12993	13092:	gap of unknown length	
13093	14661:	contig of 1569	bp in length
14662	14761:	gap of unknown length	
14762	16393:	contig of 1432	bp in length
16194	16293:	gap of unknown length	
16294	17436:	contig of 1143	bp in length
17437	17536:	gap of unknown length	
17537	19260:	contig of 1724	bp in length
19261	19360:	gap of unknown length	
19361	20445:	contig of 1085	bp in length
20446	20545:	gap of unknown length	
20546	21723:	contig of 1178	bp in length
21724	21823:	gap of unknown length	
21824	23020:	contig of 1197	bp in length
23021	23120:	gap of unknown length	
23121	24542:	contig of 1422	bp in length
24543	24642:	gap of unknown length	
24643	25843:	contig of 1201	bp in length
25844	25943:	gap of unknown length	
25944	27226:	contig of 1283	bp in length
27227	27326:	gap of unknown length	
27327	28615:	contig of 1289	bp in length
28616	28715:	gap of unknown length	
28716	29979:	contig of 1264	bp in length
29980	30079:	gap of unknown length	
30080	31099:	contig of 1020	bp in length
31100	31199:	gap of unknown length	
31200	32518:	contig of 1319	bp in length
32519	32618:	gap of unknown length	
32619	34279:	contig of 1661	bp in length
34280	34379:	gap of unknown length	
34380	35990:	contig of 1611	bp in length
35991	36090:	gap of unknown length	
36091	37493:	contig of 1403	bp in length
37494	37593:	gap of unknown length	
37594	38593:	contig of 1000	bp in length
38594	38693:	gap of unknown length	
38694	40040:	contig of 1347	bp in length
40041	40140:	gap of unknown length	
40141	41756:	contig of 1616	bp in length
41757	41856:	gap of unknown length	
41857	42923:	contig of 1067	bp in length
42924	43023:	gap of unknown length	
43024	44099:	contig of 1076	bp in length
44100	44199:	gap of unknown length	
44200	45227:	contig of 1028	bp in length
45228	45327:	gap of unknown length	
45328	46715:	contig of 1388	bp in length
46716	46815:	gap of unknown length	
46816	48171:	contig of 1356	bp in length
48172	48271:	gap of unknown length	
48272	49566:	contig of 1295	bp in length
49567	49666:	gap of unknown length	
49667	50707:	contig of 1041	bp in length
50708	50807:	gap of unknown length	
50808	51935:	contig of 1128	bp in length
51936	52035:	gap of unknown length	
52036	53211:	contig of 1176	bp in length
53212	53311:	gap of unknown length	
53312	54354:	contig of 1043	bp in length
54355	54454:	gap of unknown length	
54455	55934:	contig of 1480	bp in length
55935	56034:	gap of unknown length	
56035	57144:	contig of 1110	bp in length
57145	57244:	gap of unknown length	
57245	58469:	contig of 1225	bp in length
58470	58569:	gap of unknown length	
58570	59564:	contig of 1095	bp in length

59665	59764:	gap of	unknown length
59765	60772:	contig of 1008 bp in length	
60773	60872:	gap of unknown length	
60873	62454:	contig of 1582 bp in length	
62455	62554:	gap of unknown length	
62555	63722:	contig of 1168 bp in length	
63723	63822:	gap of unknown length	
63823	64911:	contig of 1089 bp in length	
64912	65011:	gap of unknown length	
65013	66128:	contig of 1117 bp in length	
66129	66228:	gap of unknown length	
66229	67551:	contig of 1323 bp in length	
67552	67651:	gap of unknown length	
67653	68714:	contig of 1063 bp in length	
68715	68814:	gap of unknown length	
68815	70111:	contig of 1297 bp in length	
70112	70211:	gap of unknown length	
70213	71527:	contig of 1316 bp in length	
71528	71627:	gap of unknown length	
71628	72792:	contig of 1165 bp in length	
72793	72892:	gap of unknown length	
72893	74568:	contig of 1676 bp in length	
74569	74668:	gap of unknown length	
74669	75806:	contig of 1138 bp in length	
75807	75906:	gap of unknown length	
75907	77157:	contig of 1251 bp in length	
77158	77257:	gap of unknown length	
77258	78598:	contig of 1341 bp in length	
78599	78698:	gap of unknown length	
78699	80340:	contig of 1642 bp in length	
80341	80440:	gap of unknown length	
80441	81513:	contig of 1079 bp in length	
81520	81619:	gap of unknown length	
81620	82679:	contig of 1060 bp in length	
82680	82779:	gap of unknown length	
82780	83860:	contig of 1081 bp in length	
83861	83960:	gap of unknown length	
83961	85391:	contig of 1431 bp in length	
85392	85491:	gap of unknown length	
85492	86558:	contig of 1067 bp in length	
86559	86658:	gap of unknown length	
86659	88171:	contig of 1513 bp in length	
88172	88271:	gap of unknown length	
88272	89793:	contig of 1522 bp in length	
89794	89893:	gap of unknown length	
89894	91089:	contig of 1196 bp in length	
91090	91189:	gap of unknown length	
91190	92973:	contig of 1786 bp in length	
92976	93075:	gap of unknown length	
93076	94311:	contig of 1236 bp in length	
94312	94411:	gap of unknown length	
94412	95916:	contig of 1505 bp in length	
95917	96016:	gap of unknown length	
96017	97403:	contig of 1387 bp in length	
97404	97503:	gap of unknown length	
97504	98725:	contig of 1222 bp in length	
98726	98825:	gap of unknown length	
98826	99845:	contig of 1020 bp in length	
99846	99945:	gap of unknown length	
99946	101840:	contig of 1795 bp in length	
101841	101840:	gap of unknown length	

Query Match 3.4%; Score 23; DB 2; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 23; Conservative 0; Mismatches 0; Indels 0

QY 133 CTGTTCAATAAACCCCTACGATGT 155  
|||||  
Db 25741 CTGTTCAATAAACCCCTACGATGT 25719

RESULT 6  
AP002555/C

LOCUS	AP002555	222605 bp	DNA	linear	BCT 07-MAR-2001	
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 6/20.					
ACCESSION	AP002555	BA000007				
VERSION	AP002555.1	GI:13360886				
KEYWORDS						
SOURCE	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)					
DNA						
ORGANISM	Escherichia coli O157:H7					
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.					
AUTHORS	1 (sites)					
TITLE	Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tateuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.					
	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7					
	derived from the Sakai outbreak					
	Genes Genet. Syst. 74 (5), 227-239 (1999)					
	20198780					
JOURNAL MEDLINE REFERENCE	2 (sites)					
AUTHORS	Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.					
TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655					
	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)					
	20557356					
	JOURNAL MEDLINE REFERENCE	3 (sites)				
	AUTHORS	Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.				
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak					
	Gene 258 (1-2), 127-139 (2000)					
	20564182					
	JOURNAL MEDLINE REFERENCE	4 (sites)				
	AUTHORS	Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.				
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12					
	DNA Res. 8 (1), 11-22 (2001)					
	21156231					
	JOURNAL MEDLINE REFERENCE	5 (bases 1 to 222605)				
	AUTHORS	Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.				
TITLE	Direct Submission					
	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)					
	genome project.					
	COMMENT FEATURES	Location/Qualifiers				
	source	1. .222605				
	/organism="Escherichia coli O157:H7"					
	/strain="O157:H7"					
	/sub_strain="RIMD 0509952"					
	/db_xref="taxon:83334"					
gene	74. .607					
	/gene="ECs1423"					
CDS	74. .607					
	/gene="ECs1423"					
	/note="similar to B1045_ECOLI gi 1787283 percent identity 100 in 177 aa (Conserved in E.coli K-12)"					
	/codon_start=1					
	/evidence=not_experimental					
	/transl_table=1					
	/product="putative polyprotein"					
	/protein_id="BAB34846.1"					
	/db_xref="GI:13360887"					
	/translation="MKTRIHHVQGDITKLAVDVIYNAANPSLMGGGVGDAIHRAGAALLDACLKVQQGDCPTGHAVITLAGDLPKAVVHTVGPVRRGGEQNEDELQLQDAYLNSRLVAANSYTSVAPFAISTGVYGPAAAAEIAVKTVEFITRHALPEQYIFVCYDEENAHLYERLLTQQGDE"					
	549. .2030					
	/gene="ECs1424"					
	549. .2030					
	/gene="ECs1424"					
	/note="similar to YMDC_ECOLI gi 1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"					
	/codon_start=1					
	/evidence=not_experimental					
	/transl_table=1					
	/product="putative synthase"					
	/protein_id="BAB34847.1"					
	/db_xref="GI:13360888"					
	/translation="MMKKTPTSTKDSLPNKEMNDLPRLASAVLPLCSOHPGCGCLFPLKESLDAAARYRLAEMSEHTLDVQYIWODDMSGRLFSALLAAAKRGVYRVRLILLDDNNTPGDDILRLDLSHPRIEVLNPFSEFLRLPLGTFITDESRLNRHMHKNSFTVDGVVTLVGGGRNGDAYGAGEEPLFSDLDVMAIGPVVEDVADDFARYWCKSVSPQQVLDVPEGEMADRIELPASWHDAMTHRYLRKMESPVEINHLVDGTLPLIWAQTRLLSDDPKGEGKARHSLPQRLFDIMGSPSERIDIISYFVPTAGVALLRMRVKGASLAIILTSLAANDVAVVHAGYARWKKLLRYGVELYELKPTREOSTLHDRTGTGNSGASLHAKTFSIDKTYFSGNFDPDPRSTLLNTEMGFVIESETLAQLIDKRFIQSOYDAANQWLRLDRWGRINWDRHAKKEIVLKREPAISFWKRVWRVRLASILPVEWLL"					
	complement(2038. .3195)					
	/gene="ECs1425"					
	complement(2038. .3195)					
	/gene="ECs1425"					
	/note="similar to MDOC_ECOLI gi 3025226sp p75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"					
	/codon_start=1					
	/evidence=not_experimental					
	/transl_table=1					
	/product="glucans biosynthesis protein"					
	/protein_id="BAB34848.1"					
	/db_xref="GI:13360889"					
	/translation="MNPVPAQREYFLDSTRAMLMLGIPFHISLYSSHTWVNSAEP SWLTLFNDFIHSFRMQVEFVSGYFSYMLFRLYPLKWKWVVERVGIPMLTAIPLLTLPQFTIMQYVKGKAEKSWPGLSDYKYNFTLAWELISHLWLLVLMVTLCVMIFKRI RNLNSDKTKNKSVMKLSVIFCLGICGYAVIRRTIFIVYPPILSNGTFNFTVMOTL FYLPFILLGALAFIPHLKALFTTPSRGCTLAALAFAYLLNQRYSGSDAMWYETES VITWVLGMWNVVFSFGHRLNFQSAVITYFNASLFYLVHHPHLPFLFFGATITPHI TSNWLGFLGGLIFVVGIAIILYIHLRPLLLKFLFGKPVVRENDKAPAR"					
	3588. .5123					
	/gene="ECs1426"					
	3588. .5123					
	/gene="ECs1426"					
	/note="similar to MDOG_ECOLI gi 1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"					
	/codon_start=1					
	/evidence=not_experimental					
	/transl_table=1					
	/product="periplasmic glucans biosynthesis protein"					
	/protein_id="BAB34849.1"					
	/db_xref="GI:13360890"					
	/translation="MMKMRWLSAAVMLTLYTSSSWAFSIDDVAKQAOKASLAGKGYEAPK SNLPSVFRDMKYADYQQIOFNHDKAYWNMLKTPFKLEFYHQGWYFDTVPKINIEVTATA VNRKISPDYFDGVQHDQDVKDGLFAGKVLXPINSKDNDEIVSMLGASYRVI KRIKISPDYFDGVQHDQDVKDGLFAGKVLXPINSKDNDEIVSMLGASYRVI KAGOVYGLSARGLAIDTALPSGEEPRKFEWLERPKPTDKRLTIYALLDSPRATGAY KFYVMPGRDVTVDVQSKTYLRDKVKGLGVAPLTSNLFQPNQSPANNTRPELHDSNG LSIHAGNGEWIWRPLNPKHLAVSSFMENPQGLLQGRDRFSREDDLDYLRPS AWPTPKGEMKGSVEELIPTNDETINDNIYAVMTDQLPEPKMKMFKYITITSRDED KHTPADNANWQOTRSRSTGDVGNLIRQPDGTIAFVVDFTGAEMKLPDGTIPVTAQTS IGDNGEIVESTVRYNPVTKGWRLVRVKVDKAKTKTTEMRALVNADQTLTSETWSYQLP ANE"					
	5086. .7659					
	/gene="ECs1427"					
	5086. .7659					
	/evidence=not_experimental					

ORGANISM	Salmonella typhimurium LT2
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
AUTHORS	1 (bases 1 to 24578)
	McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W., Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F., Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A., Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W., Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium LT2
JOURNAL	Nature 413 (6858), 852-856 (2001)
MEDLINE	21534948
PUBMED	11677609
REFERENCE	2 (bases 1 to 24578)
AUTHORS	The Salmonella typhimurium Genome Sequencing Project.
TITLE	Direct Submission
JOURNAL	Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT	COMMENT Supported by NIH grant 5U 01 AI43283
	Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/
	EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/
	The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.
FEATURES	Location/Qualifiers
source	1. 24578
	/organism="Salmonella typhimurium LT2"
	/strain="LT2; SGSC 1412; ATCC 700720"
	/db_xref="ATCC:700720"
	/db_xref="taxon:99287"
	/note="LT2"
	complement(87. .1472)
gene	/gene="purB"
	/note="STM1232"
	complement(87. .1457)
	/gene="purB"
	/EC_number="4.3.2.2"
	/note="similar to E. coli adenylosuccinate lyase (AAC74215.1); Blastp hit to AAC74215.1 (456 aa), 94% identity in aa 1 - 456"
CDS	/codon_start=1
	/transl_table=11
	/product="adenylosuccinate lyase"
	/protein_id="AAL20161.1"
	/db_xref="GI:16419751"
	/translation="MELSSLTAVSPVDGRYGDVKSALRGIFSEYGLLKFRVQVEVRWL QKLAHAAIKVEPAFAADANGYLDTLVANFEEDAARIKTIERTNHDVRAVEYFLKE KVAAIPALSHDVSEIFHACTSEDINNLSHALMLKRTARDEILVPWROVINAVKDLATQ YRDITPLSRTHGSEPTSTIGKEMANYAWERQFROLNQVEILKINGAVGNNAHI AAYPEVDHQFSEEFVTSIGIOWNPYTTQIEPHDYIAELDFCTARFNTLIDPDQRYW GYTALNHFKQKTAGESGSTMHPKVNPIDFENSEGNLGSNAVLHHLANKLPVSRWQ RDLTSDTVLRNLGVIGYALTAYQSTLKGYSKLEVNDRDLDELHDHNEVLAEP IQTV MRRYGIKPYEKLKELTRGKRVDAGMKQFIDSLALPEAEKTRLKAMTPANYIGRAVT LVDELK"

/gene="ECs1427"	/note="similar to membrane glycosyltransferase MDOH_ECOLI g1178287 percent identity 100 in 847 aa but differs at N-ter (Conserved in E.coli K-12)"
/codon_start=1	
/evidence=not_experimental	
/transl_table=11	
/product="membrane glycosyltransferase"	
/protein_id="BAB34850.1"	
/db_xref="GI:13360891"	
/translation="MKPGATSYLPWNKTTETIDAMPIAASEKAALPKTDIRAVHQALD AERTWAREDDSPQGSVKARLEQAWPDSLADQLIKDDRGDOLKAMPKASSMFPD PRTNPVGFWRDLRGDRTVPLRLATKEEQSEOKWRTVGTIRYILIIITLAQTV VAWTKTILPQGWALINPMDVMGVDSFPMQLPYMLGTGILIFLFAVLCWVSAG FTWALMGFTLILGHRDKYSISASTVGDEPLNPEHRTALIMPICNEDVNRVFAGLRWTS ESKVATGNKHFEDVYILSDSNPDICVAQKAWMELIAEVGEGGQIFYRRRRRVKTK SGNIIDFCRRWGSQSYMYVLADASVMTGDCLCGLVRMEANPNAGIIQSSPKASGMD TLYARCOQFATRVYGPLEFAGLHFQWLGESHYWGHNAILRVKPFIEHCALAPLGGES FAGSILSHDFVEAALMRAGWGVIAYDLPGSEYEEPLPNLLDELKDRKRWCHGNLMNE RLFLVKGHPVHRVFLTGVMSYLSAPUMFMFLASTALQVVALTEPOYFLQPRQLF PVPQWRPELAIALFASMTVLLFLPKLLSILLWCKGTKEYGGFWRTLSLLLEVLFS VLLAPVRLMFHFVYVFAFLGWGVVWNSPQRDDSTSWGEAFKRHSOLLGLVWAVG MAWLDRFLFLAPITVFSLISPFVSVISSRATVGLRTRKWKLELIPREYSPQVLVD TQRELMNRORSIDGFMHAFVNPENALATAMATARHASKVLEIARDRHVEQALNE TPEKLNRRRLVLLSDPYTMARLHFRVWNSPERISWSVSIYEGIKLNPLALRKPDAAS Q"	
7832. .8059	
/gene="ECs1428"	
7832. .8059	
/gene="ECs1428"	
/note="identical to YCEK_ECOLI g11787288 (Conserved in E.coli K-12)"	
/codon_start=1	
/evidence=not_experimental	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="BAB34851.1"	
/db_xref="GI:13360892"	
/translation="MRLIVVSIMVTLTSGCGSIISRTIPGQGHGNYYPGWQWDRDS AMRYVYILDLPFLSVDFTLLPIDIHHGPYE"	
complement(8060. .8437)	
/gene="ECs1429"	
complement(8060. .8437)	
/gene="ECs1429"	
/note="similar to MSYB_ECOLI g11787289 percent identity 99 in 125 aa (Conserved in E.coli K-12)"	
/codon_start=1	
/evidence=not_experimental	
/transl_table=11	
/product="acidic protein MsyB"	
/protein_id="BAB34852.1"	
/db_xref="GI:13360893"	
/translation="MMTWYATLEEAIDAAREEFILADNPGLDAEDANVQOFAKQYVILQ DGDIMWQVEFFADGEGEGLPMLSGEAQSVFDGDDYDEIRQEWQENTLHWEDEG	
Query Match	3.4%; Score 23; DB 1; Length 222605;
Best Local Similarity	100.0%; Pred. No. 0.26;
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 133	CTGTTCAATAAACCTACGATGT 155
Db 151662	CTGTTCAATAAACCTACGATGT 151640
RESULT 7	
AE008754/c	
LOCUS	AE008754 24578 bp DNA linear BCT 31-JUL-2002
DEFINITION	Salmonella typhimurium LT2, section 58 of 220 of the complete genome.
ACCESSION	AE008754 AE006468
VERSION	AE008754.1 GI:16419750
KEYWORDS	
SOURCE	Salmonella typhimurium LT2.

gene	complement(1461. .2160) /gene="ycfc" /note="STM1233" complement(1461. .2108) /gene="ycfc" /note="similar to E. coli orf, hypothetical protein (AAC74216.1); Blastp hit to AAC74216.1 (213 aa), 94% identity in aa 1 - 213" /codon_start=1 /transl_table=11 /product="unknown membrane associated protein" /protein_id="AAL20162.1" /db_xref="GI:16419752" translation="MIVAKNYDITLALSGICQARLVQQLAHOGHCDADALHVSLSN VIDNPSTGLVFGGSEANRLGLETLILGVLNASSRQGLNAELTRYLSLVLERKLS SAKGALNTLGRINGLQRLDHFELQSDTLMSNAGIYDVISLPGRIQVTGSPAVL QSPQVQAKVRASLGIARAAVLMHQVGGRLQLMFSRHRLTTQAKILAHLTPEL" complement(1467. .1472) /gene="ycfc" /note="putative RBS for purB; RegulonDB:STMS1H001488" complement(2112. .2117) /gene="ycfc" /note="putative RBS for ycfC; RegulonDB:STMS1H001489" complement(2134. .2142) /gene="ycfc" /note="putative -10 signal for ycfC; RegulonDB:STMLTH004479" complement(2152. .2160) /gene="ycfc" /note="putative -35 signal for ycfC; RegulonDB:STMLTH004479" complement(2189. .3340) /gene="trmU" /EC_number="2.1.1.61" /note="similar to E. coli orf, hypothetical protein (AAC74217.1); Blastp hit to AAC74217.1 (383 aa), 94% identity in aa 1 - 383" /codon_start=1 /transl_table=11 /product="tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase" /protein_id="AAL20163.1" /db_xref="GI:16419753" translation="MIEYALKFNWASIPMSPEPKVIVMGSGVDSSVSAWLQQQG YQVGLPMKNEEDGEYCTAAADLADAQACDKLGIELHTVNFAAEYDNDVFEFL EYKAGTTPPDILCNKEIKFAPLEFAEDLGADYIATGHYVRRADVNGKSRLLRGL DGNKDSYFLYTLGHEQIAQSLFPVGELEKPOVRKIAEDLGLVTAKKDGTGICFGE RKFRDLGRVLPAPGSKIITVDGDEIGEHOGLMYHTLGQRKGLGIGTKDGSDEPWV VDKDENNVLIQAQGHPRLSVGLIAQQLHWVDREFFTGTLRCTYKRYRQTDIPC TINALNDRREVIFDEPVAAPVPGQSAVFYSGEVLGGGIIIEQLPLTV" complement(3344. .3349) /gene="trmU" /note="putative RBS for trmU; RegulonDB:STMS1H001490" complement(3349. .3820) /gene="ymfB" /note="STM1235" complement(3349. .3810) /gene="ymfB" /note="similar to E. coli putative phosphohydrolase (AAC74218.1); Blastp hit to AAC74218.1 (153 aa), 84% identity in aa 1 - 152" /codon_start=1 /transl_table=11 /product="putative MutT-like protein" /protein_id="AAL20164.1" /db_xref="GI:16419754" translation="MFKPHVTVACIVHAEDKFLVVEETINGKSLWNOPAGHLEADETL AQARELWETGTGAQPOFIRHWOIAPDKTFLRFLFAIKLDHTCAIDPHDNDID CCRWVSADEILNPNLRSLPLVAESVRSYQTQGRYPILSVGEFNWPFTEGVK" complement(3815. .3820)
RBS	
gene	/gene="ymfB" /note="Putative RBS for ymfB; RegulonDB:STMS1H001491" complement(3822. .4159) /gene="STM1236" complement(3822. .4151) /gene="STM1236" /codon_start=1 /transl_table=11 /product="putative periplasmic protein" /protein_id="AAL20165.1" /db_xref="GI:16419755" translation="MKRLIIAFALLITPLSVNAASFDCQQAQAADEKAICAHLTLNDKD VEMTKYQFLKGLFAMSGRGALDAQQSWLKQRCQKADVTCITFKAYNERLKQLDVIY DHIDKPL" complement(4148. .4813) /gene="ymfC" /note="STM1237" complement(4148. .4813) /gene="ymfC" /note="similar to E. coli orf, hypothetical protein (AAC74219.1); Blastp hit to AAC74219.1 (207 aa), 83% identity in aa 1 - 205" /codon_start=1 /transl_table=11 /product="putative ribosomal large subunit pseudouridine synthase" /protein_id="AAL20166.1" /db_xref="GI:16419756" translation="MRQLISSENTMOKTSFRNHVYKRFSSRQASKSRKENQPKRVVLF NKPYDLVLPQFTDEAGRTLKDFIPQGVYAAGRLDRDSEGLLVLTNDGALQARLTQPG KRTGKIYVQVVEGIPDAAALQALRTGVLNDGPTLPAGIEIVAEPPDWLPRTPIRER KNIPTSLKLVLYEGRRNQVRMTAHVGHPTLRLIRYSMGDYTLINGLDNGQWREIAQE KDR" complement(4154. .4159) /gene="STM1236" /note="putative RBS for STM1236; RegulonDB:STMS1H001492" 4740. .4754 /note="putative binding site for FruR, RegulonDB: STMS1H00184" /bound_moiety="FruR" 4829. .6235 /gene="icdA" /note="STM1238" 4829. .4837 /gene="icdA" /note="putative -35 signal for icdA; RegulonDB:STMLTH004480"
RBS	
protein_bind	Query Match 3.3%; Score 22; DB 1; Length 24578; Best Local Similarity 100.0%; Pred. No. 0.8; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
gene	Qy 71 ACCGATTCAGTTCACGACAAAGC 92       Db 4749 ACCGATTCAGTTCACGACAAAGC 4728
-35_signal	RESULT 8 AL627269/c LOCUS DEFINITION AL627269 Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome; segment 5/20. ACCESSION AL627269 AL513382 VERSION AL627269.1 GI:16502231 SOURCE ORGANISM Salmonella enterica subsp. enterica serovar Typhi. Salmonella enterica subsp. enterica serovar Typhi Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella. REFERENCE 1 (bases 1 to 254050) AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebahia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,









```
/transl_table=11
/product="putative ABC transporter, ATP-binding protein"
/protein_id="CAC92636.1"
/db_xref="GI:15981333"
/db_xref="SPTREMBL:Q8ZBK3"
/translation="MTALEITQTKYMGVQALRGIDLHVEAGDYALLGPNAGK
STTIGISLVNKYSGKVQVFGYDIDNDIVNAKROLGLVPOEFNFPEVLIQIVTQ
AGYGVTRDALAEKYLSQLDLWSKRDRAIRLSGMRKRLMIARALHEPKLLIL
DEPTAGVDIELRRSMWGFLKELNAQGTITLTTHYLEAEMLCRNIGTIONGELVENT
TMKLLSKLESETIFDLGAKSPKLEGYGLTDTSTLEVDVKRQGLNSLFSQLN
VQGVQVOSMRNKARLELEFVLVNGHEKA"
complement(4458..5003)
/gene="YPO3406"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 198.80, E-value 8.5e-56"
complement(4641..4685)
/gene="YPO3406"
/note="PS00211 ABC transporters family signature."
complement(4959..4982)
/gene="YPO3406"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."
5329..5991
/gene="YPO3407"
/note="yadF"
5329..5991
/gene="YPO3407"

Query Match      3.1%; Score 21; DB 1; Length 216050;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GCGTAATCCATTCCTACCAG 503
      |||||
DB 53484 GCGTAATCCATTCCTACCAG 53464

RESULT 11
AF415022      884 bp DNA linear PLN 10-MAR-2002
DEFINITION   Chamaebatiaria millefolium isolate 90Chamaebatiar NADP dependent
              sorbitol 6-phosphate dehydrogenase gene, partial cds.
ACCESSION    AF415022
KEYWORDS     AF415022.1 GI:19310941
SOURCE       Chamaebatiaria millefolium.
ORGANISM     Chamaebatiaria millefolium
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids I; Rosales; Rosaceae; Rosaceae incertae sedis;
              Chamaebatiaria.
REFERENCE    1 (bases 1 to 884)
              Bortiri,E., Oh,S.-H., Gao,F.-Y. and Potter,D.
              Phylogenetic analysis of Prunus (Rosaceae) as determined by
              sequences of 56pdh
              unpublished
REFERENCE    2 (bases 1 to 884)
              Bortiri,E., Oh,S.-H., Gao,F.-Y. and Potter,D.
              Direct Submission
              Submitted (29-AUG-2001) Pomology, University of California, Davis,
              One Shields Ave, Davis, CA 95616, USA
FEATURES     Location/Qualifiers
              1..884
               /organism="Chamaebatiaria millefolium"
               /isolate="90Chamaebatiar"
               /db_xref="taxon:32222"
               join(<223..442,656..>.884)
               /product="NADP dependent sorbitol 6-phosphate
               dehydrogenase"
               join(<223..442,656..>.884)
               /codon_start=2
               /product="NADP dependent sorbitol 6-phosphate
               dehydrogenase"
               /protein_id="AAL86687.1"
```

```
/db_xref="GI:19310942"
/translation="NYELFLTRDCLSYAKIKPAVSQFETHPYQRETVIRFCOKHGVC
PVATPLGGATANTKYGVSPLGDPVLNEVAKYQKVAQICLRWNIERTAVVPKS
TKYERLKENFQVLDKLEKEDIDLINTIDRKSRITLFLSLSGLDVYA"
BASE COUNT      259 a 171 c 171 g 283 t
ORIGIN
Query Match      3.0%; Score 20; DB 8; Length 884;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 GAAAAATTTACTATGTCGAC 342
      |||||
DB 203 GAAAAATTTACTATGTCGAC 222

RESULT 12
AB064593/c      7770 bp DNA linear BCT 05-JUL-2001
LOCUS         Escherichia coli icdA gene for isocitrate dehydrogenase, complete
DEFINITION    cds.
ACCESSION    AB064593
VERSION      AB064593.1 GI:14624984
KEYWORDS     Escherichia coli
SOURCE       Escherichia coli (strain:BL21(DE3)) DNA.
ORGANISM     Escherichia coli
              Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1
              Aoshima,M., Ishii,M., Yamagishi,A., Oshima,T. and Igarashi,Y.
              Sequence polymorphism around the isocitrate dehydrogenase (ICDH)
              gene of Escherichia coli BL21(DE3) and citrate accumulation of an
              ICDH defective derivative
              unpublished
              2 (bases 1 to 7770)
              Aoshima,M.
              Direct Submission
              Submitted (03-JUL-2001) Miho Aoshima, The University of Tokyo,
              Graduate School of Agricultural and Life Sciences; Yayoi 1-1-1,
              Bunkyo-ku, Tokyo 113-8657, Japan
              (E-mail: aomih@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-5143,
              Fax:81-3-5841-5272)
              Location/Qualifiers
              1..7770
               /organism="Escherichia coli"
               /strain="BL21(DE3)"
               /db_xref="taxon:562"
               1138..2388
               /gene="icdA"
               1138..2388
               /gene="icdA"
               /codon_start=1
               /transl_table=11
               /product="isocitrate dehydrogenase"
               /protein_id="BAB61874.1"
               /db_xref="GI:14624985"
               /translation="MESKVVVPAQGGKKTITLQNGKLVNPNPIIPIBGDIGVDVTPA
               MLKVVDAAVEKAYKGEKISKMEIYTCGSTQYVQDQWLPAPETLDLIRYRVAKGP
               LTPVGGGIRSLNVALRQELDLTICLRPRTYOGTSPFVKHPETLDWVIFREMSDIY
               AGIEWKADSADAEEKVIRKFLREMGVKKIRFPEHCIGIGIKPCSEEGTKRLVRAAIEYAI
               ANDRDSVTLVHGNIMKIRFEGAKDMGQYLAAREFGGELIDGGGFWLKKVKNPTGKEIV
               IKDVIADTAQIILLRPAEYDVIACTMNLNGDYISDALAAQVGGIGIAPGANIGDECAL
               FEATHGTAPKYAGQDKVNPCSIILSAEMMLRHMGWTEAADLIVKMGEGAINAKTVTYD
               FERLMGAKLLKSGEGDAIKNM"
BASE COUNT      2259 a 1574 c 1678 g 2259 t
ORIGIN
Query Match      3.0%; Score 20; DB 1; Length 7770;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 TTCAATAAACCCCTACGATGT 155
      |||||
      |||||
```



```
/function="enzyme; purine ribonucleotide biosynthesis"
/notice="residues 4 to 459 are 88.59 pct identical to
residues 1 to 456 of 456 from E. coli K12 : B1131;
residues 4 to 459 are 88.59 pct identical to
residues 1 to 456 of 456 from GenPept :
>gb|AAC55937.1|AE005329.1 (AE005329) adenylsuccinate
lyase [Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=11
/product="adenylsuccinate lyase"
/protein_id="AA085365.1"
/db_xref="GI:21958606"
/translation="MPPMELSSLTAVSPIDRGYDKVSALRPISFSEGLLKFRVQVEV
RWLQKLAACIEIVPAFDANANAYLDKIYQEFNEQDAQRIKTIERTNHDVRAVEYF
LKEVENLPAHVSEIHFACSTEDINNSHALMLQTAQDVLPAWROIIDSIKAL
AHQYRDLPPLSRTHGQPTSTIGKELANVAYRMERFROLTOVEIIGKNGAVGYN
AHIVAYPEVDWHPQSESEVTSLGNNWNPYTOLEPHDYIAELFDCVARENTLIDFR
DIWCYIALNFKQKTIAGEIGSGSTMPHKVNPIDFENSEGNLGLSNVGHMASKLPVS
RWQRLDTSVLRNIGLVGLTALIAIYQATMKGISKEVNEAHLLEQDHNWEVLAEP
QTIMRRYGIETPKYKELKELTRGRVDAAGMQAFIDGLDLPPEEKRLKAMTPANYIGR
ATTMVDELK"
complement(5936. .6568)
/gene="y1798"
complement(5936. .6568)
/gene="y1798"
/notice="residues 3 to 209 of 210 are 75.48 pct identical to
residues 1 to 208 of 213 from E. coli K12 : B1132"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085366.1"
/db_xref="GI:21958607"
/translation="MIVAKNYDITLALAGICQSARLVQOLAHEGQDNDALNTVLRG
LLQTNPSSTLAVGDTQVLKMGLETQSVLNANRQGEAAELRYTLSLMVLKLSA
SKSAMNTLGRISQDLROLAHFDELESTMSLASIYDVVSPLGRIQVTSFPAIIQ
SPLVQAKVRATLLAGIRSAVLVQVGSRLQLMFSRNLRFKQASIIAHI"
complement(6620. .7735)
/gene="y1799"
complement(6620. .7735)
/gene="y1799"
/notice="residues 7 to 367 of 371 are 87.53 pct identical to
residues 21 to 381 of 383 from E. coli K12 : B1133;
residues 1 to 371 of 371 are 100.00 pct identical to
residues 1 to 371 of 371 from GenPept : >emb|CAC90460.1|
(AJ414149) tRNA
(5-methylaminomethyl-2-thiouridylate)-methyltransferase
[versinia pestis]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085367.1"
/db_xref="GI:21958608"
/translation="MSDNSQKVVIGMGVDSVSVALYLLQQGYQVAGLFMKNWEED
DDEYCSAATDLADAQVCDKLCMELHTVNFRAEYWDNVFELFAETKAGRTPNPDIL
KNEIKFALEFAEEDLGADYIATGHYVRQVDGKSRLLGLDGNKQDSYFLYTL
HQEIASLFPVLEKPEVRIAEQLDLTAKKKDSGTGICFIFGRKFRDPLGRLYLAQ
PGTMTVDGLGVKQGLMYHTLQGRKGLIGTGKEGDDPWVWPKDLDSNTLLVLAQ
GHEHPRLMSVLVAOQLHWVDROPVTPAPRCVVKTRYRQQDICTVTPDLDDERVDYR
DDPVAAVTPQSAVFYQGEICLGGIIEQRYPLTNP"
```

Query Match 3.0%; Score 20; DB 1; Length 10473;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 CTTTATGAAGCGCGCAACCG 539  
|||||

Db 8504 CTTTATGAAGCGCGCAACCG 8485  
|||||

RESULT 14

U55370

LOCUS

DEFINITION

U55370 32403 bp DNA linear INV 23-MAY-2002  
Caenorhabditis elegans cosmid K03B4, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

U55370

U55370.1 GI:1280096

HTG.

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

Waterston,R.

Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium

Science 282 (5396), 2012-2018 (1998)

99069613

9851916

2 (bases 1 to 32403)

Du,Z.,Le,T.T. and Kemp,K.

The sequence of C. elegans cosmid K03B4

Unpublished (2001)

3 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (17-APR-1996) Robert Waterston

4 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (01-AUG-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

5 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (01-JAN-2002) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

6 (bases 1 to 32403)

Waterston,R.

Direct Submission

Submitted (23-MAY-2002) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

email: rwaterston@wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its  
analysis see:

<http://www.wormbase.org/db/seq/sequence?name=K03B4;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is F41F3, 200 bp overlap; the 3' cosmid is ZC266,  
3000 bp overlap. Actual start of this cosmid is at base position  
197 of K03B4; actual end is at 5381 of ZC266.

NOTES:

Coding sequences below are the result of integration and manual

review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE\_INDEX.html) and The C. elegans ORFome cloning project (http://worldb.dfci.harvard.edu/), similarity to other proteins from Blastx analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES  
source

Location/Qualifiers  
1. .32403  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="v"  
/clone="K03B4"  
complement(2043. .3732)  
/gene="K03B4.5"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=K03B4.5;class  
=Sequence"  
complement(join(2043. .2121,2169. .2442,2631. .2863,  
3054. .3214,3258. .3340,3558. .3732))  
/gene="K03B4.5"  
/note="weak similarity to Xenopus dompamine D2 receptor  
(SP:P24628)"  
/codon\_start=1  
/product="Hypothetical protein K03B4.5"  
/protein\_id="AA97993.1"  
/db\_xref="GI:1280097"  
/translation="MGHSDTRFRKRVDPARTFVSFNFPPLKISSENLIIPILRSYN  
LNFQVFQEDSICHTMTSTPTTISDNFAAAVLMITNCIGFFNFLLITFS  
FYDKQKTSYLNLCVFRSINNVAILFILVYIPALIIIGESIYHPMVEVLTITAMNL  
KYNFQFSLINRLVAIYFLKYNLFGLKLAHFIIYLDVRNRYVTFFENIDRYK  
DSKFMLESYKHLAYGGVFTPDGIFWSLGLTFFPFVNIETFAIRLYLKRKTQNA  
NPKDIKKNMALFFQIFQDSLFFISVAFTMKMNLIDHRYFFSFFSQTFLWQSHIVDG  
"

complement(5626. .10741)  
/gene="K03B4.6"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=K03B4.6;class  
=Sequence"  
complement(join(5626. .5916,5967. .6269,6725. .6901,  
8300. .8587,8644. .8937,8982. .9150,10475. .10593,  
10697. .10741))  
/gene="K03B4.6"  
/codon\_start=1  
/product="Hypothetical protein K03B4.6"  
/protein\_id="AA97994.1"  
/db\_xref="GI:1280098"  
/translation="MNGYEFLEQAHLISFNATLGATSDPFFLNTSAAGNPTKGTENP  
PYVEGFMVAIFRLSMILFQKTVFSLFAFATLSLESDVDCVKWICSDITANKPDG  
CIKRIKYVPDVGVEIKTFCSEIKSVCRILKTDTEKNYNTFELDDITPIITANESDC  
YDPNVFKOCLVQGRNALOMLVDPYENKYNDAKHAKISKQCOIVKVLNTHCTPF  
DSVDFIVGGCVILNKQTSFMDCLSKLRKTPNLSKYTCENFQTKNITILKRIPT  
ADKMCWETMRKHGESSVDFQKNAATNCYSIKQSEVKEFKESRDLCDCKDQFVI  
GGVPCNIEFFKQIYNANYNECLNDELFTNLTTRREAFTFGKQCVLKLWESQEC  
PEAVQITENYLIQVLYLTIKSNDQODVSHVHKHLCDLIDDEYNKEYERLWEGDKV  
EDTGLFMIOGLRELTECESHNSGGIEIGNLKEELAMEIYGRNYSWKLYDIRRP  
SMKTLMEKILLCRLSDGDDQECRRKFVEKICGNGNATGCTETPPVTEYTTDD"  
18524. .19556  
/gene="K03B4.4"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=K03B4.4;class  
=Sequence"

join(18524. .18578,19300. .19472,19521. .19556)  
/gene="K03B4.4"  
/codon\_start=1  
/product="Hypothetical protein K03B4.4"  
/protein\_id="AA97995.1"

/db\_xref="GI:1280099"  
/translation="MRFLAAALLLVAVLLATFDVTSQSRIDRSAAWFKPLPNVRSPQ  
YXSRIRGSGSRLIPRRMYRSLPQNDYQGMKDLIDTDAQMDAEN"  
21710. .22411  
/gene="K03B4.3"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=K03B4.3a;clas  
s=Sequence"  
join(21710. .21784,21827. .21894,21959. .22223,22280. .22411)  
/gene="K03B4.3"  
/note="coded for by the following C. elegans cDNAs:  
yk602f4.3, yk602f4.5"  
/codon\_start=1  
/product="Hypothetical protein K03B4.3b"  
/protein\_id="AAK82906.1"  
/db\_xref="GI:15055382"  
/translation="MNDQEYEPSSSTESVLMPPALPOYFORPAAAPQVYSTLEPSV  
QNLRSLSHLKPLTGNOQFVOKTLENQKNPSQDDTHEFINOLADYPTTPDSTVTLHFL  
KSAGVDGSDPRVTRMISLAAKHVSDIILDAMTSARMKGLGOTKGTGKDTKYTLTEEL  
LDEILKEGHQONTTRPPYHT"  
join(21710. .21784,21836. .21894,21959. .22223,22280. .22411)  
/gene="K03B4.3"  
/note="similar to human TATA-binding protein associated  
factor 30 kDa subunit (PIR:A54981); coded for by the  
following C. elegans cDNAs: yk70h11.5, yk331g8.3,  
yk331g8.5"  
/codon\_start=1  
/product="Hypothetical protein K03B4.3a"  
/protein\_id="AA97996.1"  
/db\_xref="GI:1280100"  
/translation="MNDQEYEPSSSTESVLMPPALPORPAAAPQVYSTLEPSVONL  
RSLHLKPLTGNOQFVOKTLENQKNPSQDDTHEFINOLADYPTTPDSTVTLHFLKSA  
YGVGSDPRVTRMISLAAKHVSDIILDAMTSARMKGLGOTKGTGKDTKYTLTEELDE  
ILKEGHQONTTRPPYHT"  
22603. .23470  
/gene="K03B4.2"  
/note="for a graphical representation of this gene see:  
http://www.wormbase.org/db/seq/sequence?name=K03B4.2;class  
=Sequence"  
join(22603. .22837,23253. .23470)  
/gene="K03B4.2"  
/note="coded for by the following C. elegans cDNAs:  
yk99d5.5, yk113d4.5, yk117d8.5, yk158b11.5, yk172h12.5,  
yk113d4.3, yk99d5.3, yk117d8.3, yk158b11.3, yk172h12.3"  
/codon\_start=1  
/product="Hypothetical protein K03B4.2"  
/protein\_id="AAA97997.1"  
/db\_xref="GI:1280101"  
/translation="MPMPFTRDDYDSDPEVOAEKEMAKKDPIGFIDSSSLTSMKKA  
AAAAKDQVSTTPPNLVIGSNGAGTATVPRKSODEIVKKOKTAGQSKALKKQSR  
GISGCSRRKYSGYDDDEDEFGDDEDYDHYGGEVEDEEDTFVAKK"  
23936. .26161

Query Match 3.0%; Score 20; DB 3; Length 32403;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 41 CGAAACTCTTTTAGAAAA 60  
|||||  
Db 12095 CGAAACTCTTTTAGAAAA 12114  
  
RESULT 15  
AC106775/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-322K12, complete sequence.  
AC106775  
ACCESSION  
VERSION AC106775.3 GI:19848351  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.



```
* 2897 4278: contig of 1382 bp in length
* 4279 gap of unknown length
* 4379 contig of 1249 bp in length
* 5627 gap of unknown length
* 5628 7521: contig of 1794 bp in length
* 7522 7621: gap of unknown length
* 7622 8878: contig of 1257 bp in length
* 8879 gap of unknown length
* 8979 11131: contig of 2153 bp in length
* 11132 11231: gap of unknown length
* 11232 14216: contig of 2985 bp in length
* 14217 14316: gap of unknown length
* 14317 16088: contig of 1772 bp in length
* 16089 16188: gap of unknown length
* 16189 17922: contig of 1734 bp in length
* 17923 18022: gap of unknown length
* 18023 20155: contig of 2133 bp in length
* 20156 20255: gap of unknown length
* 20256 23371: contig of 3116 bp in length
* 23372 23471: gap of unknown length
* 23472 26665: contig of 3194 bp in length
* 26666 26765: gap of unknown length
* 26766 28982: contig of 2217 bp in length
* 28983 29082: gap of unknown length
* 29083 32399: contig of 3317 bp in length
* 32400 32499: gap of unknown length
* 32500 35642: contig of 3143 bp in length
* 35643 35742: gap of unknown length
* 35743 38529: contig of 2787 bp in length
* 38530 38629: gap of unknown length
* 38630 41401: contig of 2772 bp in length
* 41402 41501: gap of unknown length
* 41502 47514: contig of 6013 bp in length
* 47515 47614: gap of unknown length
* 47615 50863: contig of 3249 bp in length
* 50864 50963: gap of unknown length
* 50964 56189: contig of 5226 bp in length
* 56190 56289: gap of unknown length
* 56290 61081: contig of 4792 bp in length
* 61082 61181: gap of unknown length
* 61182 68596: contig of 7415 bp in length
* 68597 68696: gap of unknown length
* 68697 75882: contig of 7186 bp in length
* 75883 75982: gap of unknown length
* 75983 80902: contig of 4920 bp in length
* 80903 81002: gap of unknown length
* 81003 88297: contig of 7294 bp in length
* 88297 88396: gap of unknown length
* 88397 96759: contig of 8363 bp in length
* 96760 96859: gap of unknown length
* 96860 103626: contig of 6767 bp in length
* 103627 103726: gap of unknown length
* 103727 111301: contig of 7575 bp in length
* 111302 111401: gap of unknown length
* 111402 123190: contig of 11789 bp in length
* 123191 123290: gap of unknown length
* 123291 152037: contig of 28747 bp in length.
FEATURES
    Location/Qualifiers
        1..152037
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="CH230-347L16"
BASE COUNT 41109 a 30445 c 29440 g 43496 t 7547 others
ORIGIN
```

Query Match 3.0%; Score 20; DB 2; Length 152037;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TGAATAATACCATGACGAAAA 46

|||||
Db 39182 TGAATAATACCATGACGAAAA 39163

```
RESULT 17
AC004021
LOCUS AC004021 171370 bp DNA linear PRI 04-FEB-2000
DEFINITION Human PAC clone RP1-186K10 from 5q31, complete sequence.
ACCESSION AC004021
VERSION AC004021.1 GI:2795824
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 171370)
AUTHORS Kalicki,J., Kellen,J. and O'Brien,D.
TITLE The sequence of H. sapiens PAC clone RP1-186K10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171370)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 3 (bases 1 to 171370)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 171370)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
----- Genom Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_DJ0186K10
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

This clone was provided for sequencing by Dr. Michelle M. Le Beau in the Department of Medicine, University of Chicago, Chicago IL; and Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this clone, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics, 6:84-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).

VECTOR: pCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RP1-186K10; actual end is at 171370 of RP1-186K10. The orientation of this clone is unknown.

This clone contains STS AFM350yb1 (NID:g1051891) and EST53547 (NID:g1347444).

FEATURES

Source

Location/Qualifiers

1. 171370

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/map="5q31"

/clone="RP1-186K10"

/clone\_lib="RPC1-1"

1. 1157

/rpt\_family="L1"

1290. 2612

/rpt\_family="L1"

2617. 2679

/rpt\_family="MER81"

3511. 3643

/rpt\_family="MER1\_type"

3791. 4239

/rpt\_family="L2"

4249. 4551

/rpt\_family="MaLR"

4553. 4936

/rpt\_family="Retroviral"

5017. 5328

/rpt\_family="MER2\_type"

5361. 5408

/rpt\_family="Retroviral"

5409. 5542

/rpt\_family="MaLR"

5546. 5893

/rpt\_family="L2"

5920. 5994

/rpt\_family="MaLR"

6030. 6172

/rpt\_family="L2"

6193. 6236

/rpt\_family="L2"

6299. 6504

/rpt\_family="L2"

6505. 6536

/rpt\_family="(TGAA)n"

7620. 8444

/rpt\_family="L1"

8919. 8998

/rpt\_family="Retroviral"

9041. 10574

/note="CpG\_island (5GC-73.7, o/e=0.70, #CpGs=115)"

10335. 10420

/rpt\_family="Alu"

10416. 10523

/rpt\_family="Alu"

10535. 10594

/rpt\_family="Alu"

10732. 10988

/rpt\_family="L1"

11023. 14892

/rpt\_family="L1"

15585. 15760

/rpt\_family="MER1\_type"

18900. 18902

/rpt\_family="MIR"

18903. 18949

/rpt\_family="L2"

18950. 18972

/rpt\_family="MIR"

19618. 19768

/rpt\_family="L1"

19769. 20895

/rpt\_family="L1"

20863. 22033

/rpt\_family="L1"

22147. 24390

.

.

repeat\_region /rpt\_family="L1"  
24614. 25397  
repeat\_region /rpt\_family="L1"  
25396. 26577  
repeat\_region /rpt\_family="L1"  
27204. 27824  
repeat\_region /rpt\_family="L1"  
28298. 28456  
repeat\_region /rpt\_family="Alu"  
28460. 28510  
repeat\_region /rpt\_family="7SLRNA"  
28548. 28834  
repeat\_region /rpt\_family="Alu"  
28855. 29167  
repeat\_region /rpt\_family="Alu"  
29768. 30055  
repeat\_region /rpt\_family="L1"  
31184. 31477  
repeat\_region /rpt\_family="Alu"  
32124. 32489  
repeat\_region /rpt\_family="MaLR"  
32847. 32733  
repeat\_region /rpt\_family="MIR"  
33532. 33594  
repeat\_region /rpt\_family="L2"  
33613. 33706  
repeat\_region /rpt\_family="L1"  
34315. 34609  
repeat\_region /rpt\_family="Alu"  
34782. 34932  
repeat\_region /rpt\_family="L2"  
34935. 35152  
repeat\_region /rpt\_family="MIR"  
36512. 36802  
repeat\_region /rpt\_family="Alu"  
36882. 37126  
repeat\_region /rpt\_family="L2"  
37957. 38046  
repeat\_region /rpt\_family="L2"  
38742. 38863  
repeat\_region /rpt\_family="MIR"  
40308. 40599  
repeat\_region /rpt\_family="Alu"  
41127. 41200  
repeat\_region /rpt\_family="MaLR"  
41356. 41728  
repeat\_region /rpt\_family="MaLR"  
41917. 42152  
repeat\_region /rpt\_family="L2"  
42169. 42459  
repeat\_region /rpt\_family="Alu"  
42468. 42623  
repeat\_region /rpt\_family="L2"  
43252. 43548  
repeat\_region /rpt\_family="Alu"  
43535. 43594  
repeat\_region /rpt\_family="(TAA)n"

Query Match 3.0%; Score 20; DB 9; Length 171370;

Best Local Similarity 100.0%; Pred.No.15;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 CAGGTCAGGCGCATGACTGC 560

Db 110782 CAGGTCAGGCGCATGACTGC 110801

RESULT 18

AC116356/c

LOCUS AC116356

DEFINITION Homo sapiens chromosome 5 clone RP11-654L21, WORKING DRAFT

SEQUENCE, 3 unordered pieces.

ACCESSION AC116356

175838 bp DNA linear

HTG 10-APR-2002

VERSION KEYWORDS SOURCE ORGANISM	AC116356.2 GI:20127977 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN. Homo sapiens. Homo sapiens
REFERENCE AUTHORS TITLE JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. DOE Joint Genome Institute. 1 (bases 1 to 175838) Sequencing of Human Chromosome 5 Unpublished
REFERENCE AUTHORS TITLE JOURNAL	DOE Joint Genome Institute. 2 (bases 1 to 175838) Direct Submission Submitted (27-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA 3 (bases 1 to 175838) DOE Joint Genome Institute. Direct Submission
REFERENCE AUTHORS TITLE JOURNAL	Submitted (10-APR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA On Apr 10, 2002 this sequence version replaced gi:19745038. -----Genome Center Center: Joint Genome Institute Center Code: JGI Web site: http://www.jgi.doe.gov -----
COMMENT	Project Information Center Project Name: 1545925 Center clone name: RPCI-11_654L21 ----- Summary Statistics Consensus quality: 173992 bases at least Q40 Consensus quality: 174716 bases at least Q30 Consensus quality: 175005 bases at least Q20 Estimated insert size: 175000; agarose-fp estimation Estimated insert size: 175638; sum-of-contigs estimation Quality coverage: 5.82 in Q20 bases; agarose-fp estimation Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation. * NOTE: This is a 'working draft' sequence. It currently * consists of 3 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 21920: contig of 21920 bp in length * 21921 22020: gap of unknown length * 22021 53261: contig of 31241 bp in length * 53262 53361: gap of unknown length * 53362 175838: contig of 122477 bp in length. Location/Qualifiers 1. .175838 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="5" /clone="RP11-654L21" /clone.lib="RPCI human BAC library 11" 47733 a 36757 c 39091 g 52038 t 219 others
BASE COUNT ORIGIN	Query Match 3.0%; Score 20; DB 2; Length 175838; Best Local Similarity 100.0%; Pred. No. 15; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  Qy 541 CAGGTCAGCGCATGACTGC 560       Db 74219 CAGGTCAGCGCATGACTGC 74200  RESULT 19 AJ414149/c LOCUS
DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	Yersinia pestis strain CO92 complete genome; segment 9/20. AJ414149 AL590842 AJ414149.1 GI:15979570 Yersinia pestis. Yersinia pestis Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.
REFERENCE AUTHORS	1 (bases 1 to 193050) Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G., Prentice,M.B., Sebaihia,M., James,K.D., Churcher,C., Mungall,K.L., Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M., Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G., Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S., Karlyshev,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G. Genome sequence of Yersinia pestis, the causative agent of plague Nature 413 (6855), 523-527 (2001)
TITLE JOURNAL MEDLINE PUBMED REFERENCE	21470413 11586360 2 (bases 1 to 193050) Parkhill,J. Direct Submission Submitted (04-OCT-2001) Submitted on behalf of the Yersinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT	Notes: Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/Y_pestis/).
FEATURES source	1. .193050 Location/Qualifiers /organism="Yersinia pestis" /strain="CO92" /db_xref="taxon:632" /note="biovar: Orientalis" 372. .1151 /gene="YpO1528" 372. .1151 /gene="YpO1528" /note="Similar to Escherichia coli ferric iron reductase protein FhuF SW:PHUF_ECOLI (P39405) (262 aa) fasta scores: E(): 8.4e-33, 35.0% id in 246 aa" /codon_start=1 /transl_table=11 /product="putative ferric iron reductase" /protein_id="CAC90351.1" /db_xref="GI:15979571" /db_xref="SPTREMBL:O8ZFZ5" /translation="MSNTAQLTPSPIGLIADITELFEKTFAHFSRTLKVNADDIPEE TMSFHTWSSIDNFFTLQKYRDEYGYGNDLKPNDKALYSLWSQYFGLIIPMMLLLI EYPTIDTHHKFKVLFPHSGPEVYIQLKWSQDPTGLLERYLLNLNHHVPIAEK IESYQINGRLWNNIGYLMFWGLGFEKGLGDLYQSIINGLMELSLPLNGQDNPLY RTVMRLNGTLQRRSCCRNKLPGVRSCHDCPLKPLPLNLIS" 1303. .2850 /gene="YpO1529" 1303. .2850 /gene="YpO1529" /note="Similar to Enterobacter aerogenes L-2,4-diaminobutyrate decarboxylase ddc TR:Q9S0P8 (EMBL:AB032468) (490 aa) fasta scores: E(): 0, 39.3% id in 468 aa, and to Acinetobacter baumannii L-2,4-diaminobutyrate decarboxylase ddc SW:DDC-ACIBA (Q43908) (510 aa) fasta scores: E(): 0, 38.3% id in 431 aa" /codon_start=1 /transl_table=11 /product="putative decarboxylase" /protein_id="CAC90352.1" /db_xref="GI:15979572" /db_xref="SPTREMBL:O8ZFZ4" /translation="MYKDIYEENHFIDQGNIDITDVLCLFNKDSAEHYENVNNT ARIKAVFTLVDPKPSGVLPSELSDFYCIDLNQPLTSLDSALDELQQLYIKDAIYFHH



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 195829)  
 Moore, M.  
 Direct Submission  
 Submitted (19-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
 requests: clonerequest@sanger.ac.uk  
 On Oct 21, 2001 this sequence version replaced gi:15983876.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw:  
 SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
 database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chrl>  
 RP5-1087E8 is from the library RPCI-5 constructed by the group of  
 Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: PCYPAC2  
 This sequence is the entire insert of clone RP5-1087E8 The true  
 left end of clone RP11-1B20 is at 155425 in this sequence. The true  
 right end of clone RP11-375H24 is at 75234 in this sequence.  
 Location/Qualifiers  
 1..195829  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="q42.11-42.2"  
 /clone="RP5-1087E8"  
 /clone\_lib="RPCI-5"  
 13995..14136  
 /note="Sequence from overlapping clone RP11-375H24  
 (AL391628). Assembly confirmed by restriction digest."  
 17759  
 /note="Tandem repeat. Forced join. Gap size estimated to  
 be approximately 250bp by restriction digest data."  
 22581  
 /note="Tandem repeat. Forced join. Gap size estimated to  
 be approximately 600bps by restriction digest data."  
 38361  
 complement(38872..39056)  
 /note="Sequence from overlapping clone RP11-375H24  
 (AL391628). Assembly confirmed by restriction digest."  
 39840..39867  
 /note="Single clone region. Assembly confirmed by  
 restriction digest data."  
 195130..195532  
 /note="Sequence from overlapping clone RP11-1B20  
 (AL451047). Assembly confirmed by restriction digest."  
 BASE COUNT 57449 a 40631 c 39653 g 58096 t  
 ORIGIN

Query Match 3.0%; Score 20; DB 9; Length 195829;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 325 AAAATTTACTATGTCAGGT 344

Db 150492 AAAATTTACTATGTCAGGT 150473  
 |||||

Search completed: January 21, 2003, 00:05:22  
 Job time : 3147 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 20, 2003, 23:02:09 ; Search time 230 Seconds  
(without alignments)  
6550.378 Million cell updates/sec

Title: US-09-252-691C-1394  
Perfect score: 669  
Sequence: 1 gctataatgcgcaatgat.....ttacccttaagagaactga 669

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 20

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	23	3.4	1881	23	AAS85924
C 2	20	3.0	522	24	ABQ91613
C 3	20	3.0	603	24	ABQ91612

DNA encoding novel  
M. capsulatus gene  
M. capsulatus gene

#### ALIGNMENTS

RESULT 1

AAS85924/C  
ID AAS85924 standard; cDNA; 1881 BP.  
XX  
AC AAS85924;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #21728.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG21737.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 21728; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1881 BP; 419 A; 469 C; 533 G; 460 T; 0 other;

Query Match 3.4%; Score 23; DB 23; Length 1881;  
Best Local Similarity 100.0%; Pred. No. 0.074;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 133 CTGTTCAATAAACCCCTACCATGT 155  
|||||  
Db 591 CTGTTCAATAAACCCCTACCATGT 569

RESULT 2  
ABQ91613  
ID ABQ91613 standard; DNA; 522 BP.

```
XX ABQ91613;
AC
XX 01-OCT-2002 (first entry)
DT
XX M. capsulatus gene #1598 for DNA array.
DE
XX Micro array; gene; ds; differential expression; gene expression.
XX
XX Methyllococcus capsulatus.
OS
XX WO200255655-A2.
XX
XX 18-JUL-2002.
PD
XX
XX 14-JAN-2002; 2002WO-NO00019.
PF
XX
XX 12-JAN-2001; 2001NO-0000235.
PR
XX 12-JAN-2001; 2001NO-0000239.
PR
XX M. capsulatus gene #1598 for DNA array.
DE
XX Micro array; gene; ds; differential expression; gene expression.
XX
XX Methyllococcus capsulatus.
OS
XX WO200255655-A2.
XX
XX 18-JUL-2002.
PD
XX
XX 14-JAN-2002; 2002WO-NO00019.
PF
XX
XX 12-JAN-2001; 2001NO-0000235.
PR
XX 12-JAN-2001; 2001NO-0000239.
PR
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX (TIGR-) TIGR.
PA
XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX WPI; 2002-557818/59.
DR
XX
XX Novel DNA array useful for determining differential expression of
XX Methyllococcus capsulatus genes, comprises polynucleotides or
XX oligonucleotides representative for a selective number of Methyllococcus
XX capsulatus genes -
XX
XX Claim 14; Page 609; 678pp; English.
PS
XX
XX The invention relates to a novel DNA array giving a representation of a
XX number of Methyllococcus capsulatus genes. The method of the invention is
XX useful for determination of the differential expression of the genes of
XX M. capsulatus, and for studying gene expression on a genomic scale and in
XX gene expression assays of M. capsulatus genes. The sequences shown in
XX ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX invention.
XX
XX Sequence 522 BP; 107 A; 152 C; 145 G; 118 T; 0 other;
SQ
Query Match 3.0%; Score 20; DB 24; Length 522;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 245 GCGACAGCGAGGGGCTGCTG 264
Db 383 GCGACAGCGAGGGGCTGCTG 402
|||||
RESULT 3
ABQ91612/c
ID ABQ91612 standard; DNA; 603 BP.
XX
XX ABQ91612;
AC
XX 01-OCT-2002 (first entry)
DT
XX
XX M. capsulatus gene #1597 for DNA array.
DE
XX Micro array; gene; ds; differential expression; gene expression.
XX
XX Methyllococcus capsulatus.
OS
XX WO200255655-A2.
XX
XX 18-JUL-2002.
PD
XX
```

```
PF 14-JAN-2002; 2002WO-NO00019.
XX
XX 12-JAN-2001; 2001NO-0000235.
PR
XX 12-JAN-2001; 2001NO-0000239.
XX
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX (TIGR-) TIGR.
PA
XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX WPI; 2002-557818/59.
DR
XX
XX Novel DNA array useful for determining differential expression of
XX Methyllococcus capsulatus genes, comprises polynucleotides or
XX oligonucleotides representative for a selective number of Methyllococcus
XX capsulatus genes -
XX
XX Claim 14; Page 608-609; 678pp; English.
PS
XX
XX The invention relates to a novel DNA array giving a representation of a
XX number of Methyllococcus capsulatus genes. The method of the invention is
XX useful for determination of the differential expression of the genes of
XX M. capsulatus, and for studying gene expression on a genomic scale and in
XX gene expression assays of M. capsulatus genes. The sequences shown in
XX ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX invention.
XX
XX Sequence 603 BP; 135 A; 169 C; 180 G; 119 T; 0 other;
SQ
Query Match 3.0%; Score 20; DB 24; Length 603;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 245 GCGACAGCGAGGGGCTGCTG 264
Db 233 GCGACAGCGAGGGGCTGCTG 214
|||||
Search completed: January 20, 2003, 23:12:43
Job time : 231 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:07:39 ; Search time 50 Seconds  
(without alignments)  
4103.335 Million cell updates/sec

Title: US-09-252-691C-1394  
Perfect score: 669  
Sequence: 1 gctataatggcgcaatgat.....ttaccccttaaggagaactga 669

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB ID	Description
-----			

No matches found

Search completed: January 21, 2003, 00:30:41  
Job time : 50 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:01:24 ; Search time 56 Seconds  
(without alignments)  
5326.534 Million cell updates/sec

Title: US-09-252-691C-1394  
Perfect score: 669  
Sequence: 1 gctataatgcggcaatgat.....ttacccttaaggagaactga 669

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 393868 seqs, 222934149 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Listing first 1000 summaries

- Database : Published\_Applications\_NA:\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

No matches found

Search completed: January 20, 2003, 23:08:37  
Job time : 56 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2003, 23:07:04 ; Search time 1959 Seconds  
(without alignments)  
5530.771 Million cell updates/sec

Title: US-09-252-691C-1394  
Perfect score: 669  
Sequence: 1 gctataatgcgcaattgat.....ttacccttaaggagaactga 669

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 16154066 seqs, 8097743376 residues  
Word size : 20

Total number of hits satisfying chosen parameters: 7

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	22	3.3	497	17	AZ296025 RPCI-23-6
2	20	3.0	522	10	BE115864 UI-R-BS1-
3	20	3.0	565	10	AV523923 AV523923
4	20	3.0	699	17	BH160291 ENTSE01TR
5	20	3.0	818	17	BH570517 BOHIJ41TR
6	20	3.0	964	17	AL421622 T3 end of

c 7 20 3.0 1075 13 BI249598 BI249598 602996256

ALIGNMENTS

RESULT 1  
AZ296025  
LOCUS  
DEFINITION  
RPCI-23-62013.TJB RPCI-23 Mus musculus genomic clone RPCI-23-62013,  
DNA sequence.  
ACCESSION  
AZ296025  
VERSION  
AZ296025.1 GI:9537810  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 497)  
REFERENCE  
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret  
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other GSSs: RPCI-23-62013.TV RPCI-23-62013.TVB  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 62 row: 0 column: 13  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1. 497  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-62013"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
ECORI; Site\_2: ECORI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of ECORI and ECORI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
ECORI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life technologies)."  
BASE COUNT 134 a 123 c 74 g 166 t  
ORIGIN

FEATURES  
source  
Query Match 3.3%; Score 22; DB 17; Length 497;  
Best Local Similarity 100.0%; Pred. No. 0.31;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 44 AAACCTCTTTTGAACACCG 65  
|||||  
Db 24 AAACCTCTTTTGAACACCG 45  
|||||  
RESULT 2  
BE115864/c  
LOCUS  
DEFINITION  
UI-R-BS1-axx-c-03-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone

UI-R-BS1-axx-c-03-0-UI 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BE115864  
BE115864.1 GI:8508009  
EST.  
Norway rat  
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 522)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL  
MEDLINE  
COMMENT

Genome Res. 6 (9), 791-806 (1996)  
9704477

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
normalized embryo at 13 dpc library cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=yes.

FEATURES  
source

Location/Qualifiers

1..522  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BS1-axx-c-03-0-UI"  
/clone\_lib="UI-R-BS1"  
/dev\_stage="embryonic 13 dpc"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; site\_1: Not I; Site\_2: Eco RI; The UI-R-BS1  
library is derived from 13 dpc whole embryo tissue. For a  
detailed description of the library from which this clone  
was derived, please visit our web site at  
rategen.eng.uiowa.edu.  
TAG\_LIB=UI-R-BS1  
TAG\_TISSUE=embryo at 13 dpc  
TAG\_SEQ=ATCC"

BASE COUNT 141 a 106 c 126 g 149 t

Query Match 3.0%; Score 20; DB 10; Length 522;

Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 537 CCGACAGTCAGCGCATGA 556

Db 298 CCGACAGTCAGCGCATGA 279

RESULT 3  
AV523923

LOCUS 565 bp mRNA linear EST 01-SEP-2000  
DEFINITION AV523923 Arabidopsis thaliana aboveground organs two to six-week  
old Arabidopsis thaliana cDNA clone AP2L44a05F 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AV523923.1 GI:8603451  
EST.  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 565)  
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
of 12,028 non-redundant expressed sequence tags from normalized and  
size-selected cDNA libraries

JOURNAL  
MEDLINE  
COMMENT

DNA Res. 7, 175-180 (2000)  
20363093

Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizuekazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES  
source

Location/Qualifiers  
1..565  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="AP2L44a05F"  
/clone\_lib="Arabidopsis thaliana aboveground organs two to  
six-week old"  
/tissue\_type="aboveground organs"  
/dev\_stage="two to six-week old"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 153 a 125 c 128 g 159 t

ORIGIN

Query Match 3.0%; Score 20; DB 10; Length 565;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 588 CATTCGCTACGCCATGGCA 607

Db 543 CATTCGCTACGCCATGGCA 562

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

/note="Vector: pHOs1; Site\_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barelil, Oxford University Press, 1999)."

BASE COUNT 253 a 104 c 91 g 251 t  
 ORIGIN  
 Query Match 3.0%; Score 20; DB 17; Length 699;  
 Best Local Similarity 100.0%; Pred. No. 5.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 AACTTCTTTTAGAAACACC 64  
 |||||

Db 136 AACTTCTTTTAGAAACACC 117

RESULT 5  
 BH570517/c  
 LOCUS BH570517 818 bp DNA linear GSS 14-DEC-2001  
 DEFINITION BOHIJ41TR BOHI Brassica oleracea genomic clone BOHIJ41, DNA sequence.

ACCESSION BH570517  
 VERSION BH570517.1 GI:17822356

KEYWORDS GSS

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

REFERENCE  
 AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
 TITLE Whole genome shotgun sequencing of Brassica oleracea  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: BOHIJ41TF  
 Contact: Chris Town

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 1..818  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHIJ41"  
 /clone\_lib="BOHI"  
 /note="Vector: pHOs1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOs1 using BstXI linkers"

BASE COUNT 276 a 127 c 111 g 304 t  
 ORIGIN  
 Query Match 3.0%; Score 20; DB 17; Length 818;  
 Best Local Similarity 100.0%; Pred. No. 5.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 CCATGACGAAACTTCTTTT 54  
 |||||

Db 662 CCATGACGAAACTTCTTTT 643

RESULT 6  
 CNS06YXS  
 LOCUS CNS06YXS 964 bp DNA linear GSS 06-JUL-2001  
 DEFINITION T3 end of clone AY0AA014A03 of library AY0AA from strain CBS 6340 of Kluyveromyces thermotolerans, genomic survey sequence.  
 ACCESSION AL421622  
 VERSION AL421622.1 GI:12204820  
 KEYWORDS GSS.  
 SOURCE Kluyveromyces thermotolerans.  
 ORGANISM Kluyveromyces thermotolerans.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 REFERENCE  
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Boitot,Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekali,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
 FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL 20584711  
 MEDLINE 11152876  
 PUBMED  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 964)  
 2 (bases 1 to 964)  
 Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.  
 Genomic exploration of the hemiascomycetous yeasts: 10.  
 Kluyveromyces thermotolerans  
 FEBS Lett. 487 (1), 61-65 (2000)

JOURNAL 20584720  
 MEDLINE 11152885  
 PUBMED  
 REFERENCE  
 AUTHORS  
 3 (bases 1 to 964)  
 Genoscope.  
 Direct Submission  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 3 (bases 1 to 964)  
 Genoscope.

COMMENT  
 This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES  
 source

1..964  
 /organism="Kluyveromyces thermotolerans"  
 /strain="CBS 6340"  
 /db\_xref="taxon:4916"  
 /clone="AY0AA014A03"  
 /clone\_lib="AY0AA"  
 /note="end : T3"  
 <281..>354  
 /note="putative V (GTA) tRNA gene"

tRNA

/evidence=not\_experimental  
 <662..>820  
 /note="similar to Saccharomyces cerevisiae ORF YDR022c [CIS1 : involved in suppression of CIK1 mutation ]"

misc\_feature

BASE COUNT 267 a 240 c 234 g 222 t 1 others  
 ORIGIN

Query Match 3.0%; Score 20; DB 17; Length 964;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 CAGGTTGAAGCGGAGCCGGA 359

|||||

Db 605 CAGGTTGAAGCGGAGCCGGA 624

```
RESULT 7
LOCUS      BI249598
DEFINITION 60296256f1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5151996 5',
            mRNA linear EST 17-JUL-2001
ACCESSION  BI249598
VERSION    BI249598
KEYWORDS   BI249598.1 GI:14797131
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1075)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM11375 row: p column: 13
            High quality sequence stop: 421.
FEATURES   Location/Qualifiers
            1..1075
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="IMAGE:5151996"
                /clone_lib="NCI_CGAP_Mam5"
                /tissue_type="tumor, gross tissue"
                /dev_stage="7 months"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
                Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Robin Humphreys,
                NIH"
BASE COUNT 249 a 293 c 370 g 162 t 1 others
ORIGIN
Query Match 3.08; Score 20; DB 13; Length 1075;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 195 GAAGGACTTTATCCCGTAC 214
    |
Db 591 GAAGGACTTTATCCCGTAC 572

Search completed: January 21, 2003, 00:29:39
Job time : 1961 secs
```